PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C12Q 1/70, C12P 19/34, C12N 7/00, 7/04, 7/02, 15/00, A61K 39/12, 39/21, 39/29

A1

(11) International Publication Number: WO 99/23107

(43) International Publication Date: 14 May 1999 (14.05.99)

(21) International Application Number: PCT/US98/23107

(22) International Filing Date: 30 October 1998 (30.10.98)

(30) Priority Data:
08/962,236 31 October 1997 (31.10.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/962,236 (CIP)
Filed on 31 October 1997 (31.10.97)

(71) Applicant (*for all designated States except US*): MAXYGEN, INCORPORATED [US/US]; 3410 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Altos, CA (US). PHILLIP, Patten [US/US]; 261 La Cuesta Drive, Menlo Park, CA 94028-7532 (US). SOONG, Nay, Wei [SG/US]; 925 S. Wolfe Road #77, Sunnyvale, CA 94086 (US).

(74) Agents: QUINE, Jonathan, A. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, S...n Francisco, CA 94111-3834 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING

(57) Abstract

The invention relates to a method and compositions for modifying a phenotype of a virus, such as viral tropism and host range, by iterative sequence recombination of variant viruses and selection of improved variants.

MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of 08/962,197 filed 10-31-97. The present application claims benefit of the 08/962,197 application, which is incorporated herein by reference in its entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was partially made with federal support, NIST-ATP grant # 97-01-0240. The government may have some rights in the present invention.

FIELD OF THE INVENTION

The invention relates to methods and compositions for forced evolution of a virus genome, such as a genome of an HIV-1 virus strain, to produce a variant virus having an altered phenotype that provides a desired property that may be advantageous for development of small animal models of viral diseases, and for the development of novel therapeutic approaches to viral diseases, among others (e.g., evolving a virus to replicate in an advantageous tissue culture system). The invention relates to novel viral genomes and virions which are capable of replication in non-human animals and cells, and further relates to transgenic non-human animals and cell lines capable of supporting replication of such evolved virus variants. The invention also relates to methods for identifying novel antiviral agents.

BACKGROUND OF THE INVENTION

HIV-1 AND AIDS

Human immunodeficiency virus type I (HIV-1) is a human retrovirus that is believed to be an etiologic agent of acquired immune deficiency syndrome (AIDS), an infectious disease characterized by a profound loss of immune system function. An aspect of HIV-1 disease is the typically delayed onset of disease symptoms, such as opportunistic infections, Kaposi's sarcoma, dementia, and wasting syndrome. Often it may take 10 to 15 years after initial infection before symptoms are evident; however, in

Given this degree of diversity, it is widely believed that a vaccine based on a single strain or subtype of HIV-1 will be unsuccessful against the larger spectrum of globally circulating HIV-1 variants, as well as against new variants which continually arise. Furthermore, the HIV-1 virus appears to undergo sequence variation and functional mutation in patients; isolates from different phases of HIV-1 infection exhibit stage-specific replication characteristics (Asjo et al. (1986) Lancet 2: 660; Cheng-Meyer et al. (1988) Science 240: 80; Fenyo et al. (1988) J. Virol. 62: 4414; Tersmette (1989) J. Virol. 63: 2118).

In view of the propensity of HIV-1 to undergo rapid mutation and generate variants that are resistant to chemotheraputic agents and candidate "universal" vaccines, it is desirable to have non-human animal models of HIV-1 replication and disease in order to speed the identification and development of new generations of antiviral agents that can be used to treat resistant HIV-1 variants, or to prevent the generation of such variants in vivo. Unfortunately, such non-human models of HIV-1 disease are presently lacking.

NON-HUMAN MODELS OF HIV-1 DISEASE

The absence of a suitable animal model has remained one of the major barriers to the development of an effective therapy for HIV-1 infection. Ideally, a readily available small animal model that could sustain HIV-1 infection and develop clinical symptoms that reflect the disease in humans would prove useful for modeling pathogenesis and developing new antiviral agents. An animal model that could duplicate human immune responses would greatly facilitate the development of vaccines. Unfortunately, no current model fulfills these varied needs (for review see, Klotman et al. (1995) AIDS 9: 313; Chang et al. (1996) Transfus. Sci. 17: 89; and Bonyhadi ML and Kaneshima H (June, 1997) Molec. Med. Today pp. 246-253; Mosier DE (Sept.,1996) Hosp. Prac. Pp. 41-60).

In general, non-human animals are not susceptible to infection with HIV-1 (Morrow et al. (1987) J. Gen. Virol. 68: 2253). However, several animal models exist in which to study retroviruses related to HIV-1 and their related pathology; these include SIV in macaque monkeys, FIV in cats, and murine acquired immunodeficiency syndrome virus (MAIDS) in mice, among others. HIV-1 replicates weakly in chimpanzees, but causes no detectable disease symptoms, and chimpanzees are quite expensive and not

Ruprecht et al. (1992) AIDS Res. Hum. Retroviruses 8: 997). However, these SCID mice models produced certain results which were anomalous, such as when infected with non-cytopathic macrophage-tropic (in humans) HIV isolates the mice underwent a rapid depletion of CD4+ cells, but when infected with cytopathic, T cell-tropic HIV isolates the CD4+ cells were not depleted, the exact opposite of what occurs in the human.

Thus, the art continues to search for improved models of HIV disease using small animal models and different (i.e., non-HIV) viruses. The absence of a suitable animal model has remained one of the major barriers to the development of an effective therapy for HIV-1 infection. It is apparent from the foregoing that a need exists in the art for an improved model of HIV-1 infection to further the development of anti-HIV therapies and prophylactic agents.

Significant improvements to and new opportunities for anti-HIV therapies and antiviral screening methods could be realized if better models of HIV-1 replication and pathogenesis were available. The present invention meets these and other needs and provides such improvements and opportunities.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. All publications cited are incorporated herein by reference, whether specifically noted as such or not.

SUMMARY OF THE INVENTION

The present invention relates to methods for generating viral genotypes encoding at least one modified viral tropic phenotype, such as infectivity, virulence, and pathogenesis in a cell type, tissue, or host animal species (commonly host range; defined herein as a subset of viral tropism). The tropic phenotype modification can either permit or restrict viral infection, replication, and/or cytopatic effect in a predetermined cell type and/or host species (e.g., a non-human mammal). A basic format of the method, termed viral genome shuffling, in broad application, consists of: (1) contacting a cell strain, cell line, or non-human animal (or explanted organ therefrom), which does not naturally support substantial replication of an predetermined virus, with at least one initial infectious virion or replicable genome of said predetermined virus under replication conditions, (2) recovering a plurality of replicated genome copies of said predetermined

genome(s) produced thereby are easily distinguishable from naturally occurring viral genomes by virtue of their atypical modified viral tropic phenotype(s) which is/are normally not present in the population of naturally occurring viral genomes.

In a variation of the basic method, one or more portions of the viral genome are separately optimized or improved for function in the predetermined cell type and/or host species as distinct genetic elements isolated from the remainder of the viral genome. The optimized or improved portions of the viral genome are then either introduced into the initial viral genome(s) for use in the method, or are shuffled in by recombination with the replicated genome copies recovered after a round of replication in the host cell or host animal. In a variation, the optimized or improved portions of the viral genome can be used in conjunction with one or more heterologous polynucleotide sequence(s), such as non-viral genes or replicons to confer a desired functional or structural property, such as transcriptional regulation or translational regulation, to the heterologous sequence(s). Optimized or improved portions of a virus genome often can be marketed as a commercial product, either alone or in combination with one or more heterologous sequences.

The invention also encompasses compositions of such shuffled viral genomes encoding at least one modified viral tropic phenotype. The compositions can include a plurality of species of shuffled viral genomes, or can represent a single purified viral genome species. Certain shuffled viral genomes encode variant viruses which possess detectable phenotypes that are not naturally occurring and which can be selected for; selected phenotypes often are characterized by desirable properties, such as modified host range as compared to wildtype virus, modified cell tropism as compared to wildtype virus, and modified immunogenicity, among other desirable properties.

The invention also encompasses screening assays and kits comprising a composition of such shuffled viral genome(s) and a cell type, tissue, or host animal species for which said shuffled viral genome(s) encode a modified viral tropism or drug resistance phenotype. In an aspect, the screening assay or kit further comprises a test agent, which is typically a small organic molecule such as a nucleoside analog or protease inhibitor with a molecular weight of less than 3,000 Daltons. In an aspect, the cell type or host animal is transgenic and expresses at least one human protein which confers, either

employs a transgenic non-human cell or animal containing at least one expression cassette which encodes and expresses at least one human HIV-1 susceptibility protein. The viral genome shuffling method using these transgenic cells and/or animals as replication media produces shuffled HIV variants which have improved tropism for infection and/or replication of the transgenic non-human cells or animals. The shuffled HIV variants may be backcrossed (e.g., by recombination) to one or more HIV isolate(s), with concomitant selection for retention of the property of improved tropism for the transgenic cells or animals, thereby retaining the minimal mutations necessary for the desired tropic phenotype while "nativizing" the remainder of the viral genome to conform with the chosen HIV isolate(s). By the use of backcrossing, it is believed possible to generate, by use of the method of the invention, HIV variants substantially corresponding to essentially any HIV clinical isolate or sequence-related category thereof (e.g., group, clade, etc.), wherein the variants possess a desired phenotypic property not naturally associated with HIV; an example of such a phenotypic property can be the capacity for substantial replication in non-human cells and non-human organisms, such as for example mouse cells and transgenic mice.

In an aspect, the methods of the invention can be used to modify the immunogenic properties of a virus (i.e., the phenotype being selected for is an immunological property). For example, a virus (or collection of virus species) can be evolved to evade a host organism immune system, such as a human or mouse immune system. Also for example, a virus (or collection of virus species) can be evolved so as to mimic one or more immunologic stages of virus evolution in vivo; e.g., the viral dynamics of HIV-1 infection of a human patient is characterized by a continual natural evolution of certain immunodominant viral epitopes so as to naturally evade the human immune system - the present invention can be used to generate HIV-1 variants which mimic one or more later immunological stages of HIV infection; such variants may serve as candidate HIV-1 vaccines, among other uses.

In an aspect, the methods of the invention can be used to modify the metabolic properties of a virus (i.e., the phenotype being selected for is a resistance to one or more chemotherapeutic agent). For example, a virus (or collection of virus species) can be evolved to rapidly model the natural development of drug resistance to anti-HIV drugs. The present invention can be used to generate HIV-1 variants which are drug

unique color (shade) represents nucleotide sequence from a distinct sequence variant (e.g., from a mutated parental sequence, from a plurality of viral isolates or clades, etc.).

Figure 6. Schematic Diagram for Construction of shuffled library.

Figure 7. Schematic diagram of Passaging of a Shuffled Library to select for CHO-Tropic virus.

Figure 8. Structure of recombinant CHO-tropic envelope showing contributions from three parents.

Figure 9. HIV-1 Evolution Decision Tree.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The term "reassembly" is used when recombination occurs between identical polynucleotide sequences.

By contrast, the term "shuffling" is used herein to indicate recombination between substantially homologous but non-identical polynucleotide sequences, in some embodiments DNA shuffling may involve crossover via nonhomologous recombination, such as via cre/lox and/or flp/frt systems and the like, such that recombination need not require substantially homologous polynucleotide sequences. Homologous and non-homologous recombination formats can be used, and, in some embodiments, can generate molecular chimeras and/or molecular hybrids of substantially dissimilar sequences. Viral recombination systems, such as template-switching and the like can also be used to generate molecular chimeras and recombined viral genomes, or portions thereof.

The term "related polynucleotides" means that regions or areas of the polynucleotides are identical and regions or areas of the polynucleotides are heterologous.

The term "chimeric polynucleotide" means that the polynucleotide comprises regions which are wild-type and regions which are mutated. It may also mean that the polynucleotide comprises wild-type regions from one polynucleotide and wild-type regions from another related polynucleotide.

occurring viruses are those viruses, including natural variants thereof, which can be found in a source in nature, including virally infected individuals.

As used herein "predetermined" means that the cell type, non-human animal, or virus may be selected at the discretion of the practitioner on the basis of a known phenotype.

As used herein, "linked" means in polynucleotide linkage (i.e., phosphodiester linkage). "Unlinked" means not linked to another polynucleotide sequence; hence, two sequences are unlinked if each sequence has a free 5' terminus and a free 3' terminus.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. A structural gene (e.g., a HSV tk gene) which is operably linked to a polynucleotide sequence corresponding to a transcriptional regulatory sequence of an endogenous gene is generally expressed in substantially the same temporal and cell type-specific pattern as is the naturally-occurring gene.

As used herein, the terms "expression cassette" refers to a polynucleotide comprising a promoter sequence and, optionally, an enhancer and/or silencer element(s), operably linked to a structural sequence, such as a cDNA sequence or genomic DNA sequence. In some embodiments, an expression cassette may also include polyadenylation site sequences to ensure polyadenylation of transcripts. When an expression cassette is transferred into a suitable host cell, the structural sequence is transcribed from the expression cassette promoter, and a translatable message is generated, either directly or following appropriate RNA splicing. Typically, an expression cassette comprises: (1) a promoter, such as an SV40 early region promoter, HSV tk promoter or phosphoglycerate kinase (pgk) promoter, or other suitable promoter

(e.g., enhancer, CCAAT box, TATA box, SP1 site, etc.) that are essential for transcription of a polynucleotide sequence that is operably linked to the transcription regulatory region.

As used herein, the term "xenogeneic" is defined in relation to a recipient viral genome, mammalian host cell, or nonhuman animal and means that an amino acid sequence or polynucleotide sequence is not encoded by or present in, respectively, the naturally-occurring genome of the recipient viral genome, mammalian host cell, or nonhuman animal. Xenogenic DNA sequences are foreign DNA sequences; for example, human APP genes or immunoglobulin genes are xenogenic with respect to murine ES cells; also, for illustration, an HSV tk gene is xenogenic with respect to an HIV-1 genome. Further, a nucleic acid sequence that has been substantially mutated (e.g., by site directed mutagenesis) is xenogenic with respect to the genome from which the sequence was originally derived, if the mutated sequence does not naturally occur in the genome.

As used herein, the term "minigene" or "minilocus" refers to a heterologous gene construct wherein one or more nonessential segments of a gene are deleted with respect to the naturally-occurring gene. Typically, deleted segments are intronic sequences of at least about 100 basepairs to several kilobases, and may span up to several tens of kilobases or more. Isolation and manipulation of large (i.e., greater than about 50 kilobases) targeting constructs is frequently difficult and may reduce the efficiency of transferring the targeting construct into a host cell. Thus, it is frequently desirable to reduce the size of a targeting construct by deleting one or more nonessential portions of the gene. Typically, intronic sequences that do not encompass essential regulatory elements may be deleted. Frequently, if convenient restriction sites bound a nonessential intronic sequence of a cloned gene sequence, a deletion of the intronic sequence may be produced by: (1) digesting the cloned DNA with the appropriate restriction enzymes, (2) separating the restriction fragments (e.g., by electrophoresis), (3) isolating the restriction fragments encompassing the essential exons and regulatory elements, and (4) ligating the isolated restriction fragments to form a minigene wherein the exons are in the same linear order as is present in the germline copy of the naturally-occurring gene. Alternate methods for producing a minigene will be apparent to those of skill in the art (e.g., ligation of partial genomic clones which encompass essential exons but which lack portions of intronic sequence). Most typically, the gene segments

In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "5'-TATAC" corresponds to a reference sequence "5'-TATAC" and is complementary to a reference sequence "5'-GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length viral gene or virus genome. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each comprise (1) a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 25 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 25 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which for comparative purposes in this manner does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best

known in the art and described in Sambrook et al. et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989), Cold Spring Harbor, NY; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057, and Dunn et al. (1988) J. Biol. Chem. 263: 10878, which are each incorporated herein by reference.

As used herein the term "replication conditions" refer to aqueous conditions wherein a virus or virus genome is capable of undergoing at least one principal step of viral replication, wherein the principal step can include: attachment of virion to host cell, entry of viral genome into host cell, uncoating of virus, polynucleotide replication (RNA transcription (plus or minus strand), reverse transcription, DNA-templated DNA polymerization, viral gene expression, encapsidation, budding, and the like. In general, conditions which result in a replication phenotype (see, infra) are replication conditions. Often, suitable replication conditions can be physiological conditions. "Physiological conditions" as used herein refers to temperature, pH, ionic strength, viscosity, and like biochemical parameters that are compatible with a viable organism, and/or that typically exist intracellularly in a viable cultured mammalian cell, particularly conditions existing in the nucleus of said mammalian cell. For example, the intranuclear or cytoplasmic conditions in a mammalian cell grown under typical laboratory culture conditions are physiological conditions. Suitable in vitro reaction conditions for in vitro transcription cocktails are generally physiological conditions, and may be exemplified by a variety of art-known nuclear extracts. In general, in vitro physiological conditions can comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45°C and 0.001-10 mM divalent cation (e.g., Mg^{++}, Ca^{++}); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and often include 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton X-100) can often be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s), metal chelators, nonionic detergents, membrane fractions, antifoam agents, and/or scintillants.

As used herein, the term "statistically significant" means a result (i.e., an assay readout) that generally is at least two standard deviations above or below the mean of at least three separate determinations of a control assay readout and/or that is statistically significant as determined by Student's t-test or other art-accepted measure of statistical significance.

The term "transcriptional modulation" is used herein to refer to the capacity to either enhance transcription or inhibit transcription of a structural sequence linked in cis; such enhancement or inhibition may be contingent on the occurrence of a specific event, such as stimulation with an inducer and/or may only be manifest in certain cell types.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents are evaluated for potential activity as antiviral agents by inclusion in screening assays described herein below.

The term "candidate agent" is used herein to refer to an agent which is identified by one or more screening method(s) of the invention as a putative antiviral agent. Some candidate antiviral agents may have therapeutic potential as drugs for human use.

As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

As used herein, the term "optimized" is used to mean substantially improved in a desired structure or function relative to an initial starting condition, not

generated via genetic engineering). As long as two sequences have a region of sequence similarity, they can generally be combined.

The method can be used to shuffle xenogeneic viral sequences into a viral genome (e.g., incorporating and evolving a gene of a first virus in the genome of a second virus so as to confer a desired phenotype to the evolved genome of the second virus). Furthermore, the method can be used to evolve a heterologous sequence (e.g., a non-naturally occurring mutant viral gene) to optimize its phenotypic expression (e.g., function) in a viral genome, and/or in a particular host cell or expression system (e.g., an expression cassette or expression replicon). Fig. 2 shows an example schematic representation of recombinatorial shuffling of a collection of viral genomes having a variety of mutations or distinct genome portions; positions of mutations are indicated by an (X), and distinct genome segments (e.g., obtained from the genomes of different virus isolates) are indicated by a open box.

In an aspect of the invention, the phenotype(s) which are selected for are the tropism and/or host range of the virus. Tropism is often defined as the cell type which can be productively infected by a virus (e.g., CD4+ T cells for HIV-1, nasopharyngeal epithelium for rhinovirus, etc.), and host range is commonly defined as the species of organism in which the virus can replicate (e.g., humans, simians, mice, rats, etc.). Both tropism and host range are believed to be restricted by the specific type(s) of proteins expressed by a cell; a cell lacking expression of a necessary protein that acts as a viral receptor may fail to support infection by the virus, similarly a virus may have evolved to use a host cell protein (e.g., polymerase) in one species (e.g., human) but not in another species (e.g., mouse). The present method can be used to create variant viruses which exhibit altered tropism or host range by employing the rapid forced evolution of shuffling to generate variant viruses that are adapted to the desired tropism or host range. As an example of this, HIV-1, which does not naturally replicate in mouse cells, can be evolved to do so by the present method. Similarly, it is believed that HIV-1, which normally does not infect human fibroblasts, can be evolved to do so by the present method. The method is general and can be employed to modify tropism and/or host range of substantially any virus suitable for recursive sequence shuffling (e.g., viruses that can be rescued as infectious virions following sequence shuffling). Fig. 3 shows a schematic portrayal of virus tropism/host range evolution by viral genome shuffling to

as a cosmid clone or lambda clone). The recovered viral genome sequences can be shuffled with other viral genome sequences and/or with one or more spiked polynucleotide specie(s) (e.g., mutation-bearing gene sequences or mutation-bearing intergenic viral genome sequences), which may include optimized components of a viral genome that have been separately optimized by shuffling (e.g., a Tat gene sequence or a tar sequence of HIV-1 which has been optimized for function in mouse cells). Optimized components typically can include expression cassettes encoding viral genes, viral transcriptional regulatory sequences, origins or replication, non-coding sequences important for replication (e.g., panhandle sequences of influenza virus genome segments), LTRs, repeat sequences, and the like. For viruses with segmented genomes, individual segments may be optimized separately by recursive sequence shuffling and selection, or a combination or all of the segments may be optimized collectively for a desired phenotype; it is also possible to combine one or more cycle(s) of individual component/segment evolution with one or more cycle(s) of collective component/segment evolution, in any order.

In an aspect of the invention, a plurality of replication defective viral genomes are shuffled and the resultant shuffled genomes are selected for the capacity to replicate in a desired cell type or host organism.

In an aspect of the invention, complementing genome portions of or complete genomes of two or more distinct virus types (e.g., HIV-1 and SIV) are shuffled and phenotype selected to generate and isolate one or more shuffled variant virus genomes that have a desired phenotype (e.g., the capacity to replicate in simian cells but retain a substantial portion of the HIV-1 genome). The resultant shufflants comprising a portion of an HIV-1 (or HIV-2) genome and a portion of an SIV genome, and having functional sequences sufficient to support replication in a host cell are termed "SHIV recombinants". Kuwata et al. (1996) AIDS 10: 1331 report chimeric viruses between SIV and various HIV-1 isolates that have biological properties similar to those of parental HIV-1. Unlike the present invention, the chimeras made by Kuwata et al. are simple recombinants of discrete genome portions of SIV and HIV-1, and are not the product of recursive sequence shuffling and selection for a desired phenotype.

(1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; PCR, eds. McPherson, Quirkes, and Taylor, IRL Press, Oxford; and U.S. Patent 4,683,202, which are incorporated herein by reference). PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; (Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86, 1173; Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87, 1874; Lomell et al. (1989) J. Clin. Chem 35, 1826; Landegren et al., (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu and Wallace, (1989) Gene 4, 560; Barringer et al. (1990) Gene 89, 117, and Sooknanan and Malek (1995) Biotechnology 13: 563-564. Improved methods of cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, Ausubel, Sambrook and Berger, all supra.

Oligonucleotides for use as probes, e.g., in in vitro amplification methods, for use as gene probes, or as shuffling targets (e.g., synthetic genes or gene segments) are typically synthesized chemically according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981), Tetrahedron Letts., 22(20):1859-1862, e.g., using an automated synthesizer, as described in Needham-VanDevanter et al. (1984) Nucleic Acids Res., 12:6159-6168. Oligonucleotides can also be custom made and ordered from a variety of commercial sources known to persons of skill.

Indeed, essentially any nucleic acid with a known sequence can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (http://www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technoloigies Inc. (Alameda, CA) and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (http://www.htibio.com), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc., and many others.

Nucleic acid sequence shuffling is a method for recursive in vitro or in vivo homologous or nonhomologous recombination of pools of nucleic acid fragments or polynucleotides (e.g., viral genomes or portions thereof). Mixtures of related nucleic acid sequences or polynucleotides are randomly or pseudorandomly fragmented, and reassembled to yield a library or mixed population of recombinant nucleic acid molecules or polynucleotides.

The present invention is directed to a method for generating a selected polynucleotide sequence (e.g., a viral genome or viral gene) or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess a desired phenotypic characteristic (e.g., encode a polypeptide, promote transcription of linked polynucleotides, bind a protein, and the like) which can be selected for, and whereby the selected polynucleotide sequences are viral genomes or genes having a desired functionality and/or conferring a desired phenotypic property to a viral genome. One method of identifying novel viral genome sequences that possess a desired structure or functional property, such as having an altered tropism or host range (e.g., a human virus capable of substantial infection and replication of a non-human host), involves the screening of a large library of recombinant viral sequences (which can be a component of a viral genome - e.g., part of a viral gene, non-coding transcriptional regulatory sequence, origin of replication, - or a complete viral genome) for individual library members which possess the desired structure or functional property conferred by the novel viral genome sequence.

In a general aspect, the invention provides a method, termed "sequence shuffling," for generating libraries of recombinant polynucleotides having a desired characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related-sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. In the method, at least two species of the related-sequence polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence

(e.g., tropism of a virus in a selected cell type). Optionally, the stringency of selection can be increased between rounds (e.g., if selecting for drug resistance, the concentration of drug in the media can be increased). Further rounds of recombination can also be performed by an analogous strategy to the first round generating further recombinant forms of the gene(s) or genome(s). Alternatively, further rounds of recombination can be performed by any of the other molecular breeding formats discussed. Eventually, a recombinant form of the gene(s) or genome(s)is generated that has fully acquired the desired property.

The method of shuffling can generate libraries of polynucleotides (viral genomes, transgene polynucleotides) encoding selectable properties, including altered tropism and/or host range, which can compose all or a part of a viral genome or host cell transgene, wherein the library is suitable for function optimization of a gene or regulatory sequence or phenotypic screening. The method comprises, e.g., (1) obtaining a first plurality of library members comprising a viral genome, viral gene, viral regulatory or replication sequence, or host cell transgene (or encoding sequence or expression cassette thereof), and obtaining from said library a polynucleotide, or copy thereof, complete or partial, of at least one selected library member having a detectable desired phenotype, optionally introducing mutations into said polynucleotide or copy(ies), and (2) pooling and fragmenting, by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, typically producing random fragments or fragment equivalents, said selected polynucleotide(s) or copies to form fragments thereof under conditions suitable for PCR amplification, performing PCR amplification and optionally mutagenesis, and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool are not present in the first plurality of selected library members, said shuffled pool composing a library of shuffled selected variant viral genome sequences or transgene sequences suitable for functional screening or phenotype screening. Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual shuffled library members having the desired functional ability or phenotype. The novel shuffled viral genomes, viral genome sequences, and transgene sequences that are identified from such libraries can be used for model non-human systems of viral

In an embodiment, first plurality of selected library members is replicated under conditions wherein retroviral template switching between at least two xenogeneic viral genomes occurs, typically involving retroviral genomes or non-retroviral genes cloned into a retroviral replication system.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity. The recombination cycles (in vitro or in vivo) can be performed in any order desired by the practitioner.

The present invention provides a method for generating libraries of viral genomes or viral genetic sequences suitable for phenotype screening, such as to generate enhanced function in a cell type and/or animal species, modify viral tropism or host range, or other desired property. The method comprises (1) obtaining a first plurality of library members comprising a viral genome polynucleotide or portion thereof, (2) pooling and fragmenting said polynucleotides or copies to form fragments thereof under conditions suitable for PCR amplification and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides comprising novel combinations of viral sequences, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool comprise viral genome sequence combinations which are not present in the first plurality of library members, said shuffled pool composing a library of viral genome sequences comprising sequence combinations suitable for phenotype screening. Optionally, the plurality of selected shuffled library members can be shuffled and screened iteratively, from 1 to about 1000 cycles or as desired until library members having a desired binding affinity are obtained. Often, from 2 to 25 cycles of recursion are performed before a sufficiently optimized shufflant (i.e., selected shuffled library member) is obtained. The degree of optimization for any particular application will vary based on the specific intended use and other considerations (e.g., time, minimization of mutational drift, etc.) that are selected by the practitioner.

The invention also provides the use of polynucleotide shuffling to shuffle a population of viral genes (e.g., capsid proteins, spike glycoproteins, polymerases, proteases, etc.) or viral genomes (e.g., adenoviruses, AAV, MoMuLV, HCV, lentiviruses, retroviruses or any other known classification) to develop enhanced viral genomes having a desired phenotypic property. In an embodiment, the invention provides a method for

a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. A particular advantageous property is an altered tropism or host range which allows a human-tropic virus to infect and replicate in a non-human host animal or non-human cell type, or an altered tropism which allows a virus to replicate in a cell line which has desirable features (e.g., a cell line that has been approved by regulatory authorities, or is conveniently cultured, or the like) or altered cell tropism in a host (e.g., adenovirus that selectively infects melanoma cells and specialized Ag-presenting cells, and the like).

Forced Evolution of Models of Viral Disease

The invention provides a means to evolve virus variants and/or host cells (or organisms) that are convenient non-human model systems for studying virus-induced pathology, virulence factors, attenuated live-viral vaccine candidates, and other aspects of viral infections, as well as providing a model system for evaluating a library of agents to identify candidate antiviral agents that could find use as prophylactic and/or therapeutic drugs for human and veterinary applications.

The methods of the invention can be used to force the evolution of a virus which has a host range or tropism that limits its infectivity and/or replication to hosts which are inconvenient to use as a model system (e.g., humans or other primates, large mammals, etc.). For example, a virus which has a host range restricted to humans can be modified by recursive sequence shuffling and selection for growth in a non-human host (organism or cell culture) to produce shuffled variants that have significantly improved capacity to infect and/or replicate and produce infectious virions in the non-human host. In instances where there is no detectable infection or replication in a non-human host, shuffling of the virus of interest with a virus of a similar taxonomic type which is known to infect and/or replicate in the non-human host may generate a population of shuffled viral genomes which population contains one or more shuffled virus genomes that can replicate, at least weakly, in the non-human host. By obtaining at least one variant shuffled genome having some level of infection and replication in the non-human host (termed a "sparkplug variant"), the population of replicated virions can be collected from

efficiently promoter transcription of HIV-1 in mouse cells) benefits from optimization for function in the non-human host; recursive sequence shuffling and selection can be used to generate optimized variants of such transgene(s). Host organisms or host cells harboring transgenes which exhibit some level of functionality (e.g., ability to be infected with and/or replicate virus) can be selected for, the transgene sequence (or portion(s) thereof) recovered, and the recovered transgene sequence then shuffled with other such recovered transgene sequences and/or intentionally mutated transgene sequences to generate a population of shuffled transgene sequences that can be used to reconstitute transgenes that can be transferred into a subsequent generation of non-human host organisms or cells for one or more further rounds of selection for virus replication and shuffling, and so on. In certain embodiments, the directed evolution of the viral variants and the directed evolution of the transgene sequences of the non-human host can be done in parallel, if desired, so as to co-evolve a virus variant/host variant combination with optimized function to support virus infectivity and/or replication (or other desired feature).

Granularity of Shuffling

The "granularity" of a shuffling event refers to the relative average density of recombination joints per unit length (e.g., per kilobase) or per recombined polynucleotide molecule (e.g., per functional viral genome). For illustration, a coarse granularity could be an average of one or less recombination joint per polynucleotide resulting from a shuffling (i.e., sequence recombination event); a coarse granularity of shuffling generates a "low crossover library" (as shown diagrammatically in Fig. 5). It is often desirable to alter the granularity of shuffling in different recursion cycles, although this is not necessary in many cases. The granularity desired can frequently be selected by the practitioner and is typically accomplished by controlling the degree of recombination in the recombination format selected (e.g., for a fragmentation/reassembly format, a high degree of fragmentation will generate a small average fragment size and hence a finer granularity; increasing the number of polynucleotide species shuffled can also be used to obtain finer granularity, among other ways apparent to those skilled in the art). The average size of segment from the parental sequence(s) represented in the library of sequence-recombined polynucleotides is denoted as the "average segment length", and may be expressed by unit length (e.g., per kilobase) or as a fraction of the parental sequence (e.g., one-quarter genome of HIV-1).

a recombination reaction in which several HIV-1 clinical isolate genomes are shuffled, and a spiking mixture composed of subgenomic sequences (e.g., mutated Tat gene sequences) are included to produce a resultant shuffled library of HIV-1 genomes having enhanced sequence diversity at the Tat locus. In some embodiments, the spiking polynucleotides are viral genome components which have been optimized separately for a desired phenotype (e.g., functionality in mouse cells) and are being shuffled into a collection of viral genomes to introduce said desired phenotype into the viral genomes.

Backcrossing

After a desired phenotype is acquired to a satisfactory extent by a selected shuffled viral genome or portion thereof, it is often desirable to remove mutations which are not essential or substantially important to retention of the desired phenotype ("superfluous mutations"). Superfluous mutations can be removed by backcrossing, which is shuffling the selected shuffled viral genome(s) with one or more parental viral genome and/or naturally-occurring viral genome(s) (or portions thereof) and selecting the resultant collection of shufflants for those species that retain the desired phenotype. By employing this method, typically in two or more recursive cycles of shuffling against parental or naturally-occurring viral genome(s) (or portions thereof) and selection for retention of the desired phenotype, it is possible to generate and isolate selected shufflants which incorporate substantially only those mutations necessary to confer the desired phenotype, whilst having the remainder of the genome (or portion thereof) consist of sequence which is substantially identical to the parental (or wild-type) sequence(s). As one example of backcrossing, an HIV-1 genome can be shuffled and selected for the capacity to substantially infect and replicate in mouse cells; the resultant selected shufflants can be backcrossed with one or more genomes of clinical isolates of HIV-1 and selected for the capacity to retain the capacity to infect and replicate in mouse cells. After several cycles of such backcrossing, the backcrossing will yield HIV-1 genome(s) which contain the mutations necessary for replication and infection of mouse cells, and will otherwise have a genomic sequence substantially identical to the genome(s) of the clinical isolate(s) of HIV-1.

Isolated components (e.g., genes, regulatory sequences, packaging sequences, replication origins, and the like) can be optimized and then backcrossed with

agent library, wherein said antiviral compound inhibits replication of a nonhuman-adapted HIV-1 in said transgenic non-human animal (e.g., mouse) or transgenic non-human cells, and (3) other uses apparent to those in the art in view of this disclosure. In some embodiments, a nonhuman-adapted HIV-1 can be used in a non-transgenic, non-human host, especially when the HIV-1 genome is introduced into the host by a non-infective mechanism (e.g., electroporation, lipofection, co-transfection, etc.) and the endpoint being studied is a replication phenotype.

Transgenes and expression vectors can be constructed by any suitable method known in the art. It is often desirable to generate coding sequences for CD4, CCR5, CXCR4, and other human accessory proteins that aid viral infectivity by either PCR or RT-PCR amplification from a suitable human cell type (e.g., a T lymphocyte population) or by ligating or amplifying a set of overlapping synthetic oligonucleotides; publicly available sequence databases and the literature can be used to select the polynucleotide sequence(s) to encode the specific protein desired, including any mutations, consensus sequence, or mutation kernel desired by the practitioner. The coding sequence(s) are operably linked to a transcriptional regulatory sequence (e.g., T cell lineage-specific promoter/enhancer) and, if desired, an origin of replication (e.g., EBV ori) for episomal replication, or one or more flanking sequences having substantial sequence identity to a host chromosomal sequence to provide for homologous recombination and targeted integration of the transgene. In an embodiment, a transgene comprises a human CD4 minigene or a substantially complete human CD4 gene. Similar transgenes comprise a CCR5 and/or CXCR4 minigene or substantially complete gene. The transgenes can use the native gene transciptional regulatory sequences, or can employ an operably linked heterologous transcriptional regulatory sequence (e.g., a mouse CD4 promoter/enhancer, a CMV promoter/enhancer, a human T cell receptor gene promoter/enhancer, and the like). Often the transgene(s) and expression vector(s) will further comprise a reporter gene or a selectable marker gene (e.g., tk, neo) in a selection cassette to facilitate identification and enrichment of cells having the construct in functional form.

A wide variety of alternative transgene constructs suitable for expressing a human CD4, CCR5, and/or CXCR4 protein in non-human cells or animals will be apparent to the skilled artisan.

type" having differentiated characteristics of multiple cell types (e.g., a CD4+ cell that is predominantly a hepatocyte). A wide variety of chimeric cell-types can be generated by the skilled artisan, both as transgenic non-human animals and as cultured cell strains or cell lines.

Transgenic host cells and/or transgenic non-human animals can support infectivity by and/or replication of a virus which does not naturally infect or replicate in the non-human animal and/or cell-type. In a broad aspect, a transgenic non-human host cell or organism is generated so as to express a xenogeneic protein, or plurality of protein species, that function(s) as a receptor for attachment or entry of a virus which has a natural host range that does not include the non-human host animal. Similarly, the transgenic non-human host cell or animal can comprise a transgene which directs expression of a receptor protein to cell types or developmental stages which do not normally express said receptor protein, thereby permitting the virus of interest to infect cell types outside the natural cell tropism of the virus.

Transgenic mice, rabbits, rats, and hamster cells with sequences from human chromosome 11 and/or 12 are especially preferred for propagation of HIV-1 and SHIV chimeric virus variants. Often, such transgenic non-human animals harbor a transgene, or multiple transgenes, encoding the expression of human CD4, human CCR5, and/or human CXCR4 in T lymphocytes or other cells. Transgenes encoding the expression of other human proteins can be similarly constructed and transgenic animals produced therefrom.

Bridge Cells and Bridge Organisms

In some cases, the desired non-human host cell or organism may be incapable of supporting replication of the virus in part because the desired host (e.g., mouse) is too distant phylogenetically from the natural host (e.g., human). The desired host may lack certain proteins necessary for replication of the virus, or may have equivalent host cell proteins which are too divergent from the natural host protein(s) in order to function effectively with the virus that has naturally evolved to function in its natural host. In such instances, it may be impossible to generate sparkplug variants by directly transferring shufflants of primary viral isolates into the divergent non-human host cells or organisms, and alternative strategies to adapt the virus to grow in the desired non-human host will need to be used.

complementing portion of a bridge viral genome, a chimeric viral genome which is capable of replication in the bridge host is created.

Most frequently, it is useful to incorporate those portions of the bridge virus genome which are believed to encode functions that are substantially distinct between the subject viral genome and the bridge viral genome. These regions can be identified by highly divergent sequences between the two viral genomes, or can be regions containing genes known or believed to be important in controlling host range (e.g., surface glycoproteins such as the env gene of a retrovirus such as HIV-1). Often such critical genes are (1) viral glycoproteins, (2) polymerases or other transcription factors which must interact with host proteins or polynucleotides, or (3) viral non-coding sequences or secondary structures which must interact with host proteins (e.g., HIV-1 TAR hairpin sequences).

For example, if the subject virus is HIV-1 and the desired bridge host is a non-human primate, it is often advantageous to incorporate portions of a simian immunodeficiency virus (SIV) viral genome to create a chimeric HIV/SIV viral genome, termed a "SHIV" viral genome. Kuwata et al. (1996) AIDS 10: 1331 describe chimeric SHIV viruses composed of gag, pol, vif, vpx, nef, and LTR from SIVmac and vpr, tat, rev, vpu, and env of various HIV isolates. Chimeric viral genomes can be created by mixing predetermined portions of each genome on the basis of intelligent prediction of their functionality in the bridge host (as per Kuwata et al.), or the chimeric viral genomes can be created by shuffling all or portions of each viral genome with the other viral genome and selecting shufflants which possess the desired phenotype, which is typically enhanced replication in the bridge host. A variation employs chimeric oligonucleotides as PCR primers, wherein the chimeric primer has a first portion complementary to a HIV sequence and a second portion complementary to a SIV sequence to generate by PCR shuffled SHIV variants wherein the recombination junctions are principally the boundaries between the HIV sequence and the SIV sequence in the chimeric primers (see, Fig. 4). In this way, recombination joint location can be biased according to the practitioner's choice, which may be random, pseudorandom, or intelligent. The present invention thus provides for a collection of shuffled chimeric viral genomes which can then be subjected to selection for a desired phenotype.

sequences, by a recursive sequence recombination method (e.g., fragmentation/reassembly format, template-switching, and the like) to produce shufflant SHIV genomes. Optionally, a mutagenic process (e.g., error-prone PCR, chemical mutagenesis, spiking with mutagenic oligonucleotides having random or pseudorandom sequence variation) is performed on the recovered SHIV genome sequences, either before, during, or after the shuffling step. The shufflants are rescued as infectious virions and a subsequent cycle of infection of the monkey cells is commenced. The cycle of (1) recovering virions or proviral DNA from cells in having a replication phenotype, (2) shuffling and optionally mutagenizing the sequences, and (3) rescuing infectious virions from shufflant genomes, is repeated until a desired level of replication in the host cells is obtained or until replication competence of the shufflants plateaus. After SHIV shufflants having the desired phenotype (e.g., improved replication in monkey cells) are obtained, they are used to infect mouse cells (e.g., mouse lymphocytes from a transgenic mouse expressing human CD4+ on peripheral T lymphocytes), and replicated virions or proviral DNA from cells having a replication phenotype are recovered so as to select for SHIV shufflants that are competent to replicate in mouse cells. The recovered SHIV genomes may then be subjected to additional round(s) of shuffling (optionally including mutagenesis) and selection to optimize replication in the mouse cells. When a desired level of replication is obtained in the mouse cells, the SHIV shufflants are backcrossed to (i.e., shuffled with) the parent HIV-1 viral genome or a collection of HIV-1 genomes, optionally including a mutagenesis process, and the resultant shufflants are rescued as infectious virions and used to infect mouse cells, and a recursive process of backcrossing to parent HIV-1 genome(s) and selection for replication in mouse cells will produce a chimeric HIV-1 viral genome that is predominantly derived from the parent HIV-1 genome and which contains a minimal degree of SIV sequences and/or mutations necessary to provide the desired level of replication in mouse cells.

Evolution of Component Sequences by Shuffling

The present method of shuffling can be used to optimize subgenomic components, such as structural genes, transcriptional regulatory regions, packaging sequences, replication sequences, subgenic functional domains, gene clusters, complete genomes, and the like), for a particular phenotype (e.g., functionality in a novel host species or cell type). The optimized components can then be shuffled into a replicable

codon usage in a host cell, to optimize translational efficiency or RNA processing efficiency, and the like). Improved function of a structural gene product is determined by a suitable assay system and selection of such assay system is dependent upon the nature of the specific gene product and can be selected by those skilled in the art. For example and not limitation, an assay to measure function of a viral gene product (e.g., pol) necessary for replication of a virus can comprise measuring the replication of a virus genome lacking said viral gene (e.g., pol) in a cell or organism in which the viral gene is encoded by and expressed from an expression cassette that is separate from the viral genome (e.g., an expression vector encoding pol); this variation can be termed "complementation in trans". A library of such expression cassettes encoding the component can be screened for functionality, and functional species selected (e.g., enriched for) , and shuffled to produce a selected, shuffled library of shuffled component sequences that can be subjected to one or more additional rounds of functional selection and/or shuffling, so as to obtain one or more sequences encoding optimized component species (termed "optimized component sequences", or with specific reference to structural genes - "optimized structural gene component sequences").

When the component to be optimized is a transcriptional regulatory sequence, the readout is typically the improved transcription of a reporter polynucleotide (or the polypeptide encoded thereby) in the particular host cell or organism selected by the practitioner. The transcription of the reporter polynucleotide sequence can be detected by a method to detect transcription (e.g., PCR, LCR, hybridization with a labeled complementary sequence polynucleotide probe, inactivation of a conditionally lethal or screenable gene product by antisense hybridization, and the like), or by a method to detect an encoded gene product of the reporter polynucleotide (e.g., luciferase, β-galactosidase, HRP, GFP, and other suitable detectable reporter proteins). However, other readouts for optimizing transcriptional regulatory sequences can also be used; for example and not limitation, the readout can be the level of expression of a viral structural gene operably linked and transcriptionally modulated by the transcriptional regulatory sequence. Improved function of a transcriptional regulatory sequence is determined by a suitable assay system and selection of such assay system is dependent upon the nature of the specific gene product and can be selected by those skilled in the art. For example and not limitation, an assay to measure function of a viral transcriptional regulatory sequence

Exemplary Components

A component can be any subgenomic sequence comprising more than 10 consecutive nucleotides of a viral genome, typically comprising all or a substantial portion of a viral structural gene, transcriptional regulatory sequence, or replication control sequence. A component can also be any nonviral sequence of more than 10 consecutive nucleotides of a structural gene or transcriptional regulatory sequence from an animal cell genome (or mRNA pool), wherein said sequence encodes a protein involved in viral entry, viral transcription, viral replication, or viral egress, or wherein said sequence regulates transcription of a viral sequence (whether as in integrated provirus or as an episomally replicating viral genome).

To illustrate the invention and not to limit it, the following non-exhaustive list of viral components can be obtained from an HIV-1 genome: gag MA (P17), gag CA (p24), gag NC (p7,p6), protease (p15), reverse transcriptase/RNase H (p66,p51), integrase, Env (gp120/gp41), Tat (p16/p14), Rev (p19), Vif (p23), Vpr (p10-15), Vpu (p16), Nef (p27/p25), Vpx (p12-16), Tev (p28),U3 sequence, U5 sequence, primer binding site sequence (PBS), polypurine tract (PPT), repeat region (R), long terminal repeat (LTR), minimal HIV promoter (NF-κB site, Sp1 sites, TATA box, transcription initiation site, Tat-responsive element (Tar), Rev-responsive element (RRE), splicing signals, and other open reading frames or transcriptional regulatory regions of the HIV genome. Similar components from SIV can often be used, and may supplement or replace the cognate components (or portions thereof) of the HIV-1 component.

To illustrate the invention and not to limit it, the following non-exhaustive list of nonviral (host cell) components can be obtained from a genome or mRNA pool of an animal cell and are believed important in HIV-1 entry, replication, or egress: cellular factors that bind to Tar or to Tat, factors encoded on human chromosome 12 that contribute to the transcriptional activity of Tat, CD4, CXCR4, CCR5, p56lck, NF-κB, Sp1, other coreceptors for HIV-1 attachment or entry, other host factors necessary for HIV-1 replication, and the like.

Although the examples provided reference HIV-1, those skilled in the art will be capable of selecting components from the particular virus type they desire to work with.

Shuffling and Selection

A plurality of species of a component are obtained, either by mutating a starting component specie to create a pool of mutated component species or by beginning with a plurality of component species (e.g., component "alleles" obtained from a plurality of virus isolates or even different virus types, such as HIV-1 and SIV), or other methods. The pool of component species can be either be incorporated into reporter constructs, introduced into host cells, and selected for a desired phenotype prior to the first round of shuffling, or may be initially shuffled before any selection is performed.

The plurality of component species is shuffled by a suitable sequence recombination method (e.g., by DNase fragmentation and PCR-based reconstitution of overlapped joints, or by any of the variety of suitable sequence shuffling methods described herein and elsewhere, and as is known in the art) to generate a library of sequence-recombined ("shuffled") component polynucleotides. The library of shuffled component sequences, typically in the form of reporter constructs, are introduced into host cells by a suitable method (e.g., transfection, electroporation, viral infection, lipofection, and the like) and the resultant pool of introduced shuffled reporter constructs are selected or screened for the desired functionality of the shuffled component sequences. Those library members (or progeny thereof) which comprise shuffled component sequences having a desired phenotype are recovered and the resultant pool of selected shuffled component sequences can be put through one or more additional cycles of recursive sequence shuffling to further optimize for the desired phenotype(s), or for additional phenotype(s). Mutagenesis and/or spiking can be used in conjunction with shuffling to further enhance the sequence diversity in one or more rounds of shuffling. Suitable mutagenesis methods are selected at the discretion of the practitioner, but for illustration and not limitation can include: site-directed mutagenesis by mutagenic oligonucleotide, error-prone PCR, chemical mutagenesis, mutagenic irradiation, propagation of polynucleotides in error-prone hosts, and the like.

Recovery of Selected Polynucleotide Sequences

A variety of selection and screening methods will be apparent to those skilled in the art, and will depend upon the particular phenotypic properties that are desired. The selected shuffled viral genome sequences can be recovered for further shuffling or for direct use by any applicable method, including but not limited to: recovery of virions from cells or extracellular medium (e.g., ascites, serum, spent cell

component is optimized for function in the context of a replicable viral genome (i.e., a context-optimized component).

Rescue of Infectious Virus from Cloned Viral Sequences

One objective of the general method of shuffling viral genome sequences to produce shuffled sequences encoding a desired phenotype ultimately is the generation of virus variants that exhibit altered host range and/or cell tropism. In order to accomplish this expeditiously, it is sometimes preferable to employ a system to rescue infectious virus from cloned viral sequences. This may often be as simple as transfecting a viral genomic polynucleotide into a suitable host cell in which the viral genome can express necessary replicative functions, replicate the genomic polynucleotide, encapsidate the genomic polynucelotide, and egress the cell (if appropriate). Sometimes it is necessary to utilize a helper cell line or helper virus to obtain replication and packaging. The helper cell line or helper virus typically provides a function in trans (e.g., viral polymerase) that facilitates an important step in viral replication and/or packaging. In the case of some viruses (e.g., negative strand RNA viruses), it can be necessary to form appropriate ribonucleoprotein complexes in order to support efficient replication of the viral genome in a cell (WO97/12032 and U.S. Patent 5,166,057, for example). The skilled practitioner will select the rescue system appropriate for the particular virus that is to be shuffled.

With regard to HIV, there are a variety of suitable methods to recover infectious molecular clones, such as integrated or circularly permuted, non-integrated proviral forms, or subgenomic proviral sequences that can be reconstituted into full-length provirus (Gibbs et al. (1994) AIDS Res Hum Retroviruses 10: 607; Ghosh et al. (1993) Virology 194: 858; Li et al. (1991) J. Virol. 65: 3973; Fredriksson et al. (1991) Virology 181: 55). PCR can be used to construct infectious molecular clones of HIV-1 from full-length provirus. Infectious molecular clones of HIV can be obtained from the NIAID AIDS Research and Reference Reagent Program (Bethesda, MD) or other publicly available source, or can be generated by the practitioner. Salminen et al. (1995) Virology 213: 80 disclose a method for recovering full-length HIV-1 provirus DNA from primary virus cultures by using PCR; this methodology can be used to recover HIV provirus from starting materials (e.g., HIV primary isolates) for subsequent shuffling, and to recover proviral DNA from selected HIV shufflants. Landau and Littman (1992) J.

recombination of HIV-1 sequences and selection for mutant, shuffled, and/or chimeric HIV-1 sequences that have enhanced function and replicability in mouse cells. There are many alternative approaches to making such murine replicable HIV viral genomes by shuffling; these alternative variations will be apparent to the practitioner, and some specific variations are described herein for illustration and not limitation.

Generation of HIV Competent to Replicate in Mouse Cells

Viral genomes from HIV isolates can be shuffled with each other, with mutated HIV genomes, and/or with SIV or murine-tropic retroviral (MLV) genomes. The shufflants can be introduced into mouse cells expressing human CD4, human CCR5, and CXCR4 and selected for capability to replicate in the mouse cells and produce infectious virus that is capable if infecting such transgenic mouse cells. Once a desired level of replication of the evolved HIV shufflants is achieved, additional properties may be selected for, such as independence from human CD4 by performing additional cycles of recursive shuffling and selection on mouse cells expressing CCR5 and/or CXCR4 and lacking human CD4. In a variation, HIV-1 genome sequences are shuffled with a HIV-2 env gene, which is independent from CD4 for viral entry, to produce shufflants that encode an env protein that does not obligatorily require human CD4 for virus entry. Such env genes may be chimeras between a HIV-1 env and a HIV-2 env, or may be predominantly or exclusively HIV-2 env sequence, and possibly include additional mutations introduced as part of the recursive shuffling process.

Backcrossing to Specific Clades or Parent HIV Isolates

HIV-1 isolates can be grouped according to phylogenetic sequence similarities into categories referred to in the art as clades (Gao et al. (1994) AIDS Research and Human Retroviruses 10: 1359). Once a murine-replicable HIV-1 shufflant having a satisfactory capacity to replicate in mouse cells is obtained, recursive sequence recombination can be used to backcross the replicable HIV variant to one or more naturally-occurring HIV sequences, such as the wild-type parental backbone(s) from which the HIV variant was derived or to other HIV isolates. By performing multiple cycles of shuffling (backcrossing to a naturally-occurring HIV sequence and/or to a consensus sequence representing one or more clades), and selection for retention of the phenotype of replication in mouse cells, it will be possible to make murine-replicable variants of essentially any HIV isolate or clade representative sequence. In order to

other host ranges and/or cell tropisms. Many non-human primate species can be used as a source of cells, which may be propagated in primary cell culture or immortalized by a variety of art-known methods. Alternatively, or in combination, with passaging the HIV virus (or mutagenized and/or shuffled variants thereof; including SHIV chimeras) in non-human primate cell cultures, it is also possible to passage these viruses in intact non-human primates, and recover the evolved virus variants from tissues or fluids of the primates and subject the recovered variants to recursive sequence shuffling and selection for replication in the non-human primate.

Virus Evolution in a Transgenic Mouse

HIV-1 shufflants can be introduced directly into transgenic mice harboring a transgene that encodes and expresses a human receptor for HIV (e.g., CD4, CCR5, CXCR4, etc.), and infective and replicable variants can be recovered from tissues (e.g., lymphoid tissues, peripheral blood lymphocytes) or fluids (e.g., serum, ascites) of the mouse. The mouse may also have a reservoir of human lymphoid tissue, such as a SCID/hu mouse with a human thymus/liver sandwich implanted under the kidney capsule. The reservoir of human lymphoid tissue can serve as a reservoir of human cells competent to replicate shuffled HIV variants such as may replicate poorly in mouse cells at early cycles of a forced evolution to modify host range to include mice. The human cell reservoir can amplify, by replication, the number of variant HIV viruses that can replicate in the mouse cells, as well as increase the background of HIV variants which are replicating solely in the human reservoir cells. However, since subsequent selections can be done with virus recovered from the animal and replicated in the absence of human cells, the increased background of human-specific HIV is not problematic.

Mixed Particle Infection (High MOI)

Superinfecting host cells at a high multiplicity of infection (MOI) can be used to advantage to increase the recombination between viral genomes. Preferably an MOI of 5 to 50 or greater is used to enhance recombination during the viral replication cycle in the cell.

Identification of Novel Human HIV Cofactors

Mouse cells non-permissive for HIV-1 infection can be used for expression screening of human cDNA libraries to identify cDNA sequences that encode proteins which confer permissivity to HIV-1 infection and/or replication. In an

chimeric protein comprising the ligand-binding domain of a steroid receptor and the site-specific recombinase), so as to produce site-specific recombination among the proviral genome sequences and thereby effect shuffling and the production of shuffled viral genome variants.

Combinations

Combinations of the shuffling and selection strategies disclosed herein can be used.

Defective HIV Variants Having Enhanced Safety

Once shuffled, HIV variants that are capable of substantial replication in mice are established, one or more viral genomic sequence(s) necessary for the altered host range and/or tropism and that function in trans can be deleted from the shuffled and adapted viral genome and provided in trans in the host cell or animal (e.g., as a transgene expression cassette), so that the host provides the helper function that complements the replication of the virus, but the resultant virus that is produced is non-infective for organisms that lack the helper function. In this way, a model system of a transgenic animal providing an internal helper function can be used in conjunction with a replication-deficient HIV virus to develop antiviral drugs and study HIV disease without fear that infectious, replication-competent virus will be produced and infect lab workers or escape into human or animal populations.

Attenuation Phenotypes

Shuffling can be used to generate virus variants having attenuated phenotypes, such as reduced pathogenicity and/or virulence. One general type of such attenuated variants are the temperature-sensitive and/or cold-adapted mutants. In this aspect, selection of shuffled variants would select for shuffled viral genomes that replicate efficiently at reduced (or elevated) temperature. Other attenuation types can also be selected for.

Other Phenotypes

The present method can be used to generate variant viruses having a wide variety of altered phenotypes. Illustrative examples not intended to limit the scope of the invention are: (1) capability to replicate in a non-permissive cell, (2) host range and/or cell tropism distinct from naturally-occurring wild-type virus, (3) improved virus titer (e.g., virulence), (4) decreased pathogenicity and capacity to produce disease, (5)

Codon Modification Shuffling

Procedures for codon modification shuffling procedures are described in detail in SHUFFLING OF CODON ALTERED GENES, Phillip A. Patten and Willem P.C. Stemmer, Attorney Docket Number 018097-028500US, filed September 29, 1998. In brief, by synthesizing nucleic acids in which the codons which encode polypeptides are altered, it is possible to access a completely different mutational cloud upon subsequent mutation of the nucleic acid. This increases the sequence diversity of the starting nucleic acids for shuffling protocols, which alters the rate and results of forced evolution procedures. Codon modification procedures can be used to modify any viral nucleic acid herein, e.g., prior to performing DNA shuffling. This can have the benefit of allowing the virus to adapt to a host cell's codon selection, e.g., prior to shuffling.

Use of RecA

The frequency of homologous recombination between nucleic acids can be increased by coating the nucleic acids with a recombinogenic protein, e.g., before or after introduction into cells. See Pati et al., *Molecular Biology of Cancer* 1, 1 (1996); Sena & Zarling, *Nature Genetics* 3, 365 (1996); Revet et al., *J. Mol. Biol.* 232, 779-791 (1993); Kowalczkowski & Zarling *in Gene Targeting* (CRC 1995), Ch. 7. The recombinogenic protein promotes homologous pairing and/or strand exchange. The best characterized *rec*A protein is from *E. coli* and is available from Pharmacia (Piscataway, NJ). In addition to the wild-type protein, a number of mutant *rec*A-like proteins have been identified (e.g., *rec*A803). Further, many organisms have *rec*A-like recombinases with strand-transfer activities (e.g., Ogawa et al., *Cold Spring Harbor Symposium on Quantitative Biology* 18, 567-576 (1993); Johnson & Symington, *Mol. Cell. Biol.* 15, 4843-4850 (1995); Fugisawa et al., *Nucl. Acids Res.* 13, 7473 (1985); Hsieh et al., *Cell* 44, 885 (1986); Hsieh et al., *J. Biol. Chem.* 264, 5089 (1989); Fishel et al., *Proc. Natl. Acad. Sci. USA* 85, 3683 (1988); Cassuto et al., *Mol. Gen. Genet.* 208, 10 (1987); Ganea et al., *Mol. Cell Biol.* 7, 3124 (1987); Moore et al., *J. Biol. Chem.* 19, 11108 (1990); Keene et al., *Nucl. Acids Res.* 12, 3057 (1984); Kimiec, *Cold Spring Harbor Symp.* 48, 675 (1984); Kimeic, *Cell* 44, 545 (1986); Kolodner et al., *Proc. Natl. Acad. Sci. USA* 84, 5560 (1987); Sugino et al., *Proc. Natl. Acad. Sci. USA* 85, 3683 (1985); Halbrook et al., *J. Biol. Chem.* 264, 21403 (1989); Eisen et al., *Proc. Natl. Acad. Sci. USA* 85, 7481 (1988); McCarthy et al., *Proc. Natl. Acad. Sci. USA* 85, 5854 (1988); Lowenhaupt et al., *J. Biol. Chem.* 264, 20568 (1989). Examples of such recombinase proteins include *rec*A,

EXPERIMENTAL EXAMPLES

The following examples are illustrative and not limiting. One of skill will realize a variety of parameters which can be changed to achieve essentially the same results.

EXAMPLE 1: EVOLUTION OF NOVEL PHENOTYPES IN HIV BY INTRA- AND INTERCLADE SHUFFLING

The diversity of HIV sequences in natural and laboratory isolates is utilized to generate a library of recombinant HIV sequences from which strains with desired characteristics are selected. These include novel tropisms on cells from species normally refractory to HIV infection, the use of alternate receptors to enter cells and improved replication kinetics.

Method

Sources of HIV Sequences

Subgenomic sequences of various regions of the HIV genome are obtained from: 1) available molecular clones of different HIV strains; 2) PCR using consensus or degenerate primers from genomic DNA of chronically infected tissue culture cells; 3) RT-PCR of HIV particles from supernatants of chronically infected cells or patient fluids.

A wide collection of such sequences are collected from multiple clades of HIV. These subgenomic HIV sequences are cloned into bacterial plasmids that will be used as templates.

DNA Shuffling

Shuffling is performed either by: 1) Directly performing circular shuffling of plasmids carrying analogous regions of the HIV genome from different isolates; or 2) Pooling PCR fragments amplified from plasmids carrying analogous regions of the HIV from different isolates genome and performing linear shuffling.

Shuffled material is amplified using primers incorporating specific restriction sites. These restriction sites enable the shuffled amplified fragments to be functionally cloned into the backbone of an infectious HIV clone (pNL4.3) containing the remainder of the HIV genome as in the case of MLV full length reconstruction where the Moloney MLV clone provided the backbone (*see* infra). This reconstitutes a full length HIV clone. A library of recombinant shuffled HIV clones are thus constructed. The library is propagated and amplified in *E. coli* to obtain DNA for transfection.

efficiently e.g., hematopoietic cells; b) non-replicating cells; c) more stable retroviral particles; d) higher titers; and e) site specific integration of delivered genes.

It also demonstrates the feasibility of building a general library from which different novel phenotypes affecting different parts of the viral life cycle can be selected.

MLV Strains

Without being exhaustive, all or subsets of MLV strains listed below are used for family shuffling.

Strain/isolate	Class
1) Friend	Eco
2) Rauscher	Eco
3) 292E Cl. 15	Eco
4) 292A	Ampho
5) Mo- Ampho	Ampho
6) Moloney	Eco
7) AKR	Xeno?MCF
8) Gross	Eco
9) Balb V2 / BC 169	Endo/Xeno
10) Balb V1/BC 194	Endo/Xeno
11) Balb V2 /BC 177	Endo/Xeno
12) AT 124	Xeno
13) NZB cl 15.	Xeno
14) AKR 13	MCF
15) AKR 247	MCFSource

Proviral DNA

The Moloney MLV was obtained as an infectious proviral clone. All the other strains were obtained as biological clones or preparations and used to infect Mus Dunni cells. Genomic DNA from infected Mus Dunni cells was used as template to PCR subgenomic fragments of the various MLV strains. These subgenomic fragments are thus cloned into bacterial Bluescript plasmids.

Shuffling

A 3 kb region encompassing the 3' 500 bp of pol, the entire env and 3' LTR constitute the subgenomic fragment to be shuffled. The DNA used for shuffling is derived from in vitro PCR using plasmid clones as templates. Mixtures of PCR amplified fragments from all or subsets of the 15 strains are used in the shuffling process. The fragments are Dnase digested into 400- 1.5 kb size range. These are reassembled by 30- 45 cycles of annealing and extension without primers. The assembled mixture is then amplified with PCR. These shuffled fragments are then subcloned into the Moloney backbone and transformed into E. coli libraries. Other subgenomic regions, e.g., the gag

Characterization of Desired MLV Clones

Recovered infectious MLV clones with the desired phenotypes are characterized by traditional techniques such as titers, Westerns for viral antigens, reverse transcriptase activity, tropism, mapping and sequencing.

Cell Lines Used

Name	Description
3T3	Murine embryo
SC-1	Feral Mice
Mus Dunni IIIc	Feral mice tail fibroblast
P19	Mouse embryonal carcinoma
F 9	Mouse embryonal carcinoma
Mv-1 Lu	Mink Lung
293	Human kidney embryonic
CHO-K1	Chinese Hamster Ovary
BHK-21	Baby Hamster kidney (Syrian)
Don CHL	Hamster
PA317	Retroviral packaging; amphotropic/ 3T3

EXAMPLE 3: CREATION OF ADENOVIRUS HOST RANGE MUTANTS BY DNA SHUFFLING

Tissue tropism is a problem that exists in most of the viral vectors. For example, retroviruses do not target specific cells and only integrate into dividing cells. Ad2 and Ad5 infect most human cells but do not infect or propagate in lymphocytes, keratinocytes, and hematological malignant cells. The host range determinants of Ad infection include viral and host factors. Cells must have Ad receptors (still unknown) and integrins in order to be permissive for Ad infection, and the Ad viruses must have appropriate fibers, penton base, and early genes in order to infect and propagate in the cells. It has been previously shown that by infecting nonpermissive cells (Vero) with a high MOI of Ad12 and continuously passaging the infected cells for many weeks, an adapted Ad12 mutant with altered host range (grows well in Vero) can be isolated. By shuffling the viral DNA, this adaptation process is facilitated, and new host range mutants

culture cells and in woodchucks. These systems will be especially useful for rapid screening and testing of new drugs.

Hepatitis B virus (HBV) infection is the major risk factor in the development of chronic hepatitis and hepatocellular carcinoma (HCC). As much as 15% of the population is chronically infected in areas where this virus is highly prevalent such as in eastern Asia and sub-Saharan Africa. A large scale epidemiology study has shown that approximately 40% of the male HBV carriers will eventually die of HCC.

None of the established cell lines is susceptible to infection of HBV derived from serum , or produced by HBV-producing cell lines. HBV can only infect primary human hepatocytes and the hepatocytes of chimpanzee. Thus, chimpanzee, an endangered species which is expensive and allows only limited experimentation, represents the only available animal model. There is a woodchuck hepatitis virus (WHV) which is homologous to HBV and causes chronic hepatitis and HCC in the woodchucks. However, the pathology of WHV infection in woodchucks is somewhat different from that of HBV in human or chimpanzee. Thus, the availability of permissive cell lines and small animal models, in which HBV can infect and propagate, would be valuable for the testing of therapeutic vaccines and drugs.

The HBV replication cycle involves multiple steps, including virus attachment and entry, formation of covalently closed circular DNA, transcription, RNA packaging and reverse transcription, (+) strand synthesis, and viral assembly and release. Many of these steps involve interactions between HBV genome/gene products and those of the host cell. Therefore, the inability of HBV to infect and replicate in culturable human cells and in woodchuck may be caused by multiple blocks, and the number of mutations required to generate a mutant capable of replication in nonpermissive cells can be large. This possibility is also suggested by the fact that, despite intensive research in this field, so far no such a host range HBV mutant has been isolated. DNA shuffling is uniquely suited to obtaining novel mutants with complex genetic compositions which require multiple combinations of mutations or existing alleles. Therefore, DNA shuffling may be a promising approach to solving the problem of evolving HBV to grow in human hepatic cell lines and in woodchucks.

viruses and particularly retroviruses are notorious for their ability to evolve their way around biological blocks. This process has been utilized in many studies to evolve viruses with new phenotypes such as expanded tropism (Vahlenkamp, T.W. et al. Journal of Virology 71, 7132-7135 (1997) Taplitz, R.A. & Coffin, J.M. Journal of Virology 71, 7814-7819 (1997), drug resistance (Balzarini, J. et al. Journal of Virology 67, 5353-5359 (1993) Dianzani, F. et al. Antiviral Chem. Chemother. 4, 329-333 (1993) and promoter activity (Barklis, e., Richard, M. & Jaenisch, R. Cell 47, 391-399 (1986)). Components of evolved viral variants, for example LTR elements have been incorporated into improved viral vectors (Robbins, p.B. et al. Journal of Virology 71, 9466-9474 (1997)).

Adaptation of viruses to new host cells typically requires prolonged passaging and selection. This is due to the necessity for the continuous generation and selection of variants before an effective solution is found. Usually this involves only a few mutations. Biological adaptations that require constellations of mutations or novel combinations of functional domains may not be achieved without long periods of replication and frequent extinctions. In this example, we demonstrate that DNA shuffling can dramatically accelerate viral evolution towards desired phenotypes by enhancing recombinatorial processes in vitro.

In DNA shuffling (e.g., .Stemmer, P.C. Nature 370, 389-391 (1994)), similar input sequences are first randomly fragmented. The fragments are then reassembled through multiple cycles of self-priming polymerase chain reaction. Because of the complementary overlapping ends, a fragment from one parental sequence can prime off a template from another parental sequence. DNA shuffling thus generates a population of recombinant sequences which is then screened or selected for improved phenotypes. The process can be applied recursively to independently selected sequences to recombine useful variations, often with synergistic effects. The diversity of the input parental sequences can be generated by mutagenic processes or, more effectively, by using several natural occurring sequences (Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W.P.C. DNA Nature 391, 288-291 (1998)) (natural diversity). DNA shuffling thus accelerates natural processes of evolution by the rapid and efficient generation of diversity through errors and recombination, followed by selection. Many single and multigene systems have been dramatically improved using this process (Patten, P.A.,

clones exhibited patterns different from any of the parents. This represents a lower limit for recombination frequency as many other nucleotide changes may not be detected. To assess the viability of the library, 5 pools of 4 clones each were transfected into 293/G1 cells. The viral supernatants were tested for the ability to transduce G418 resistance into 3T3 and Mus Dunni cells. Four of the 5 pools were able to strongly transduce G418 resistance into at least one of the cell types. Thus, if each positive pool only had one infectious clone, this would give a frequency of 20% (4/20) which represents a lower limit for the viability of the library.

Passaging of Library / Selection

Selection was performed by passaging the shuffled library supernatant on a mixture of CHO KI and Lec8 cells as illustrated in Fig 7 and described supra. A control mixture of the six unshuffled parents were passaged identically. A small proportion of Lec 8 cells was mixed in during passaging to support a low level of replication in a permissive cell type that was as similar to the target CHO K1 cells as possible. Lec 8 cells are CHO K1-derived mutants whose ecotropic receptors are believed to be more accessible because of a defect in their glycosylation pathways. This renders them permissive to infection by some ecotropic MLVs (See also Wilson, C. & Eiden, M.V.E. J. Virol. 65, 5975-5982 (1991); Miller, D.G. & Miller, D. J. Virol. 66, 78-84 (1992). Wang, H. et al. J. Virol. 70, 6884-6891 (1996)). Friend 2, Friend 9 and Moloney MLVs produced from transfected 293/G1 are able to infect Lec 8 cells fairly efficiently (Table 1).

Table 2 shows the progress of the selection for both the control unshuffled parents and the shuffled library. By titering at each stage, the changing 'infection profile' of the viral population was monitored. The initiating transfections into 293/G1 for the shuffled library produced supernatants that gave titers on 3T3, Mus Dunni and Lec8 that were on the order of 10^2 fold lower than that for the control parental pool.

The infectious activities of both the control parental pool and the shuffled library fell to similar levels after one passage on the coculture cells, even though the shuffled library started out with 10^2 fold lower titers. This indicates that the shuffled library is actually fitter than the parental pool under the coculture selection conditions. This point is underscored after a second passage of the viral pools. The parental pool essentially becomes extinct after the second passage (extremely low activity can be

(Clones 1-6, 8, 10-12) that represents the 'master sequence'. Clones 7 and 9 are slightly different from the dominant pattern but are also distinct from any of the parents.

Clones 3, 10 and 11, corresponding to the dominant pattern and the variant clone 7 were transfected into 293/G1 cells, and the supernatants were tested for infectious activity (Table 3a). Surprisingly, all of these clones had drastically diminished infectivities for CHO K1 when compared to the passage 5 pools from which they were isolated. Relative to titers on Lec 8 cells, the infectivities of these clones for CHO K1 was on the order of 10^{-5} or less, 100-1000 fold lower than that for passage 5 supernatants. This suggested that the 'CHO-tropic' clone in passage 5 was not represented by any of the four clones tested. The infectious efficiency of passage 5 supernatants on CHO K1 relative to the other cell types is about 10-3-10-2 (range from several subsequent titrations). This could be interpreted in two ways: 1) The predominant virus particle in this supernatant can infect CHO K1 at an relative efficiency of 10^{-3}-10^{-2}; 2) there is one viral particle in every 100 -1000 infectious particles that can infect CHO K1. If the latter were true, this rare clone would be expected to be selected for under our passaging regime and increase in frequency. However the CHO K1 infectious efficiency apparently has stabilized at 10^{-3}-10^{-2} suggesting the viral population has achieved some state of 'equilibrium'. This is supported by the clear dominance of one clone as shown by restriction analysis. These observations indicated that the clone that conferred CHO K1 infectivity was not missed, but that this activity was masked in our clones.

Table 2: Titers of parental and shuffled library passage supernatants on coculture cells

	Titer cells	Transfection	Coculture Passage Number 1	2	3	4	5	5B&
Control Parentals	3T3	1.2×10^7	7×10^3	0	0	0	0	ND
	MD	$3 \times 10^{5*}$	4×10^2	0	0	0	ND	ND
	Lec8	$1.4 \times 10^{4*}$	30	0	0	8	0	ND
	CHOK1	0	0	0	0	0	0	ND
Shuffled Library	3T3	3×10^5	1×10^3	16	7	10^2	2×10^4	10^6
	MD	10^4	4×10^2	20	30	45	ND	10^5
	Lec8	$9 \times 10^{2*}$	40	19	19	10^2	$>5 \times 10^4$	10^5
	CHOK1	0	0	0	0	10	5×10^2	10^3

MD : Mus Dunni
ND : not done
* from separate experiment which gave comparable titers on 3T3
& supernatants from later cultures split from passage 5

Table 3a : Diminished CHO K1 Infectivity after Growth in 293/G1 cells

CLONE #	Titer Cells 3T3	Mus Dunni	Lec8	CHO KI
3	10^4	10^6	10^6	10*
7	50	10^5	10^4	0
10	10^3	10^5	10^5	0
11	10^4	10^5	10^5	10*

* estimated from a single G418 resistant colony in the 10^{-1} titration well.

Table 3b: CHO K1 infectivity is reconstituted after passage through Lec8/G1 cells

CLONE #	Titer Cells 3T3	Mus Dunni	Lec8	CHO KI
3	10^6	10^6	10^6	10^3
7	10	10^3	10^2	0
10	10^5	ND	10^6	10^3
11	10^5	10^6	10^6	10^3

and rapid dilution of virus production and to the observed rapid decline of the infection during passaging.

Sequences of Recombinant Envelope

The envelope sequences of Clones #3 and #11 are most consistent, with a four fragment recombination between three of the Friend parents (Fig. 8). The sequence of clone #3 can be explained by recombination alone, while that for clone#11 has an additional silent base change at position 231. Nucleotide differences between the parents allow us to map the regions where crossovers took place. It is not surprising that the Moloney and 292E sequences were not included in the selected clones. Recombination events involving these two parents may be under-represented as they have lower degrees of identities with the Friend sequences. Because of their greater divergence, recombination events may also have a higher probability of generating non-viable clones. Although the 3' LTR and parts of pol were also shuffled, it is unlikely that they play significant roles in the new tropism of the recombinant clones. Pol is highly conserved between ecotropic MLVs and is not known to have a role in entry. Cloning of recombinant envelope sequences which excludes the 3' LTR, using the Sfi I site in pol and a conserved Cla I site towards the end of the envelope is sufficient to confer CHO K1 tropism (data not shown). This indicates that the changes in the LTR were not necessary.

DNA shuffling was used to improve individual genes as well as multigene pathways. In this example, we report an application of shuffling to evolve a desired phenotype in a viral system. The ability to infect CHO K1 cells was evolved by shuffling sequences from a defined set of ecotropic parental MLVs. No *a priori* assumptions were made of the changes required to overcome the CHO K1 entry block other than that the envelope was involved.

Predominantly, envelope sequences from the six parents were shuffled to generate a library of about 1 X 10^6 clones. At least one third of these were recombinant. This shuffled library consistently gave 100 fold lower titers than the parental pool upon initial transfection into 293/G1 cells. This is caused by the generation of many lethal and debilitated sequences by the shuffling process. Thus the fitness of the naïve library is lower than the unshuffled parental pool. This reflects the 'cost' of the shuffling process in generating diversity at the expense of population fitness.

mutation in the glycosylation pathway of Lec 8 cells, their golgi can only import galactose at 2% of wild type levels, resulting in low efficiency of terminal addition of galactose and sialic acid at N-linked glycosylation sites (Deutschert, S.L. & Hirschberg, C.B. Mechanism of Galactosylation in the Golgi Apparatus. J. Biol. Chem. 261, 96-100 (1996). An altered glycosylation pattern of the envelope when expressed in Lec 8 may be responsible for enhancing CHO K1 infectivity. Glycosylation patterns of retroviral envelopes produced in different CHO glycosylation mutant cell lines are clearly different (Fenouillet, E., Miquelis, R. & Drillien, R. Virology 218, 224-231 (1996). Friend 21 is more divergent than any of the other Friend parents. In the segment that Friend 21 contributes to the recombinant clones, three amino acid residues (378, 413 and 447; Fig. 8) that are specific for Friend 21 are positioned 1-3 residues away from N-linked glycosylation sites. These may influence the efficiency of sugar addition which may in turn affect the overall conformation of the envelope. Cellular processing and conformation of retroviral envelope glycoproteins are known to be heavily dependent on glycosylation signals. The receptor binding domain (Heard, J.M. & Danos, O. J. Virol. 65, 4026-4032 (1991)) of the recombinant envelope is provided by Friend 2 and Friend 9 parents, both of which can infect Lec 8 cells. It may be that this receptor binding domain in juxaposition with the altered glycosylation signals from Friend 21 is processed in Lec 8 cells to produce an envelope that is able to reinfect Lec 8 cells and to a lesser degree, to infect CHO K1 cells. The glycosylation mediated block of CHO K1 receptors can be relieved by inhibiting glycosylation in these cells. This may have the effect of making the receptors more accessible to the envelope. The same effect might also be achieved by under-glycosylating the retroviral envelope itself. This modification of retroviral tropisms by altering the glycosylation pattern of envelopes may represent a novel mechanism that has not been reported previously.

The passage of parental viruses produced from 293/G1 through Lec 8 results in poor production of infectious viruses (Friend 9) or in progeny viruses that cannot reinfect Lec 8 efficiently (Friend 2 and Moloney). This may be a direct result of the altered glycosylation pattern of these parental envelopes in Lec 8 cells. Under-glycosylation of the Friend 9 envelope may lead to gross misfolding while for Friend 2 and Moloney, this may lead to conformational changes that result in the inability of the envelope to bind the Lec8 receptor efficiently. The rapid abrogation of the parental

Methods

Cell Lines

Cell lines were obtained from American Type Culture Collection. A retroviral vector expressing the G418 resistance marker (from Gene Therapy Laboratories, University of Southern California) was introduced into these cells which were then subjected to G418 selection at 0.8-1 mg/ml. About 20-100 resistant colonies for each cell type were pooled. These G418 resistant lines are denoted with a '/G1' suffix.

Viruses

Friend MLV (ATCC VR 245) was obtained as a spleen extract containing a mixture of three viruses. An ecotropic 292E strain (ATCC VR 1326) was obtained as a supernatant from infected NIH 3T3 cells. Genomic DNA from Mus Dunni cells infected with these stocks were used to recover proviral sequences of the different MLV strains (below). Plasmid pNCA (gift from S. Goff, University of Columbia) contains a full length , non-permuted copy of the wild type Moloney MLV proviral DNA in a pBR322 based vector (Colicelli, J. & Goff, S.P. J. Mol. Biol. 199, 47-59 (1988)).

Cloning of Envelope Sequences

Genomic DNA was isolated from Mus Dunni infected with Friend or the 292 ecotropic (292 E) MLV strains using the Puregene kit (Gentra Biosystems) and manufacturer's protocols. Primers were designed to amplify Friend and 292E MLV proviral sequences based on the published Moloney MLV sequence (Genbank accession number M76668). The upstream sense primer Mol PolESn straddles the SfiI site in the pol gene which is highly conserved between ecotropic MLV strains. The downstream antisense primer, MolU5as is positioned at the 3' end of the U5 sequence. A NotI site is also included in the 5' tail of this primer(Fig. 1). PCR was performed using reagents from the GeneAmp XL PCR kit (PE Applied Biosystems). Final concentrations of Mg acetate, primers and each dNTP were 1.25 mM , 0.5 uM and 200 uM respectively. PCR fragments from the 292E and Friend amplifications were processed and eventually cloned into a modified pNCA (see below) acceptor backbone using the SfiI and NotI unique sites. Plasmid pNCA was modified by inserting a Not I site just downstream of the 3' LTR of the Moloney MLV sequence. A unique Sfi site exists in the 3' region of the pol gene. Cleavage of the modified pNCA plasmid with Not I and Sfi I excises about 0.5 kb of pol, the entire env and 3' LTR. The remaining backbone then served as an acceptor

Shuffling of Proviral Sequence and Library Construction

These six clones were used as templates for PCR amplification to generate material for shuffling. MolPolESn and pBRas, an antisense primer in the pBR322 vector sequence just downstream of primer MolU5as were used to amplify a specific 3.2 kb product. PCR products from each of the six parents were purified and mixed together in equimolar amounts. This mixture was then digested with DNAse I (Sigma). DNAsed fragments in the size range of 0.7 – 1.6 kb were purified and used in the shuffling reaction essentially as described before (Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P.C. Nature Biotechnology 14, 315-319 (1996)). The completed shuffling reactions were used as templates for preparative PCR using primers MolPolEsn and MolU5as. Products from this were purified and digested with NotI and SfiI. These fragments were then cloned into similarly digested modified pNCA acceptor backbone and transformed into XL-10 Gold competent cells (Stratagene). Approximately 1 X 106 colonies were obtained and pooled and used to prepare library plasmid DNA. Several independent colonies were also individually picked and analyzed. Fragments representing the shuffled region were amplified from these clones. These PCR fragments were digested simultaneusly with Bgl I, Cla I, Dra I, Dra III and Sac II. The digests were run out on a 1.5% agarose gel and compared to the restriction patterns of the parents. Clones were also assayed for viability.

Library Passaging / Selection for CHO K1 Tropic Virus (Fig 2)

Library plasmid DNA was transfected into 4 plates of 293/G1 cells as described above. 40 ml of supernatant was collected. About 5 ml of this was used for titering while 10 ml (polybrene was added to 8 ug/ml) was passaged onto a coculture of CHO K1/G1 (90%) and Lec 8/G1 cells (plated at a total density of 5 X 105 cells/ 100 mm plate. The coculture cells were exposed to this supernatant for 24–48 hours before being replaced with fresh F12 Ham (Gibco BRL) media with 10% FBS. When the coculture cells had grown to 90-100% confluency, fresh media was added and left on the cells for 48 hours. This supernatant was collected, filtered and used for titering and for passaging onto fresh coculture cells. As a control to account for natural recombination and adaptation, an equimolar mixture of the six parental clones were transfected, passaged and assayed identically to the library supernatant.

that either that do not express CD81 or only express it at low levels. To enhance expression of CD81 and thus facilitate infection of cells with evolved HCV variants, cell lines are optionally stably or transiently transfected with a CD81 cDNA expression vector. Cells lines that could be used in the screening, after transfection with CD81, include, but are not limited to, Hela, Cos-1, Cos-7, CHO, 293, U937, HL60, Jijoye, Jurkat, Hep G2, C3A, TF-1, Baf-3. Methods for stable transfection are known to those skilled in the art, and are described for example by van der Merwe et al. (J. Exp. Med. 185, 393-403, 1997) and Lanier et al. (J. Immunol., 154, 97-105, 1995).

Shuffling is performed on the entire genome of HCV or subgenomic portions or both. The size of the HCV genome is within the range of previous sequences that have been successfully shuffled (e.g. adenovirus, with > 20kb shuffled). Furthermore, the genome of HCV is highly heterogeneous with the assignment of at least six HCV types encompassing 11 subtypes. The most divergent HCV isolates differ from each other by more than 30% over the entire genome. Sequence identities lower than this have been successfully shuffled (e.g. Cephalosporinase). Moreover, HCV, like many RNA viruses circulates as a quasispecies, further adding to natural diversity which can be harvested for shuffling.

Protocol for Shuffling and Selection of HCV

Prepare large quantities of genomic and/or subgenomic fragments of multiple species of HCV by PCR or by amplification in bacteria. These are obtained as full length or partial molecular clones, or from clinical samples.

DNA shuffling is performed, including e.g., DNAse I digestion , PCR assembly, (e.g., a long range, high-fidelity PCR protocol). The PCR can be performed such that a promoter such as T7 is incorporated at the 5' end. PCR fragments (full length or subgenomic) are optionally cloned into a HCV genomic cDNA template with a promoter incorporated to reconstitute full length molecular clones. Runoff transcription is performed to generate libraries of potentially infectious transcripts. Pools of RNA transcripts are transfected into target cells. As noted above, target cells include those which express CD81, either naturally, or following transfection with a CD81 coding nucleic acid. Infectious sequences are recovered by PCR, e.g., from virions or negative strain (replicated) RNA by RT-PCR. It is also possible to enrich or select for replicating

If a virus is evolved or engineered to replicate in murine cells, it will have many mutations relative to wild type HIV-1 which may be unnecessary for replication in murine cells and which will compromise it as a valid model for AIDS. DNA shuffling provides a solution to this problem because one can backcross a mutant of interest with wild type strains. This natural feature of shuffling technology is used to perform *in vitro* backcrosses of evolved variants with wild type HIV-1 strains of commercial interest. This step will ensure that only those mutations necessary for viral propagation in the mouse are preserved, thereby optimizing the predictive value of this laboratory model for the human disease. These evolved viruses will be used in conjunction with the double transgenic mice to identify novel small molecule drugs and prophylactic and treatment vaccines.

The experimental strategy is schematized in Figure 9. HIV-1 is adapted to grow in murine tissue culture cells using both "top down" and "bottom up" approaches. These mutants are further evolved to replicate in hCD4+, hCCR5+ double transgenic mice, and to cause pathogenesis. These mutant HIV-1 isolates are backcrossed to wild type HIV-1 isolates to obtain a virus that can replicate in the transgenic model while being maximally similar to wild type human HIV-1 isolates. Figure 9 schematizes the strategic choice tree that used to prioritize objectives and to decide when to move on to subsequent modules of HIV shuffling and design.

<u>Top down approach</u>

In the top down approach, a mutant virus is identified that can replicate, however weakly, on hCD4+ hCCR5+ murine cells. This is done by testing existing HIV-1 isolates and by constructing libraries of novel HIV-1 recombinants using DNA shuffling. Initial selection is performed in tissue culture cells. Weakly replicating viruses serve as starting points for further evolution. To increase the efficiency of selecting a mutant virus that can be propagated in murine cells, DNA shuffling is used to recombine the diversity that exists in the natural HIV population. Libraries of novel recombinants are generated containing mutants that are capable of replicating in the hCD4+ hCCR5+ murine target cells. Viral replication is quantitated by measuring p24 production and viral reverse transcriptase activity. The goal is to evolve a virus that yields a tissue culture infectious dose-50 (TCID-50) of 1-10% the level produced by wild type HIV-1 on human cells. This approach initially yields weakly replicating virus.

obtained for GFP reporter construct driven by tat, tar and the HIV LTR. Second, mouse cell expression of a GFP reporter gene encoded at the 3' end of the HIV-1 LTR is obtained that is 10 - 50% of the level expressed by a wild type HIV-1 GFP reporter virus integrated into human cells. Third, viral titers of 1 - 50% of wild type HIV-1, as quantitated by p24 concentration or quantitative RT-PCR measurements of viral RNA in the supernatant in standard spreading infection assays is obtained. The kinetics of growth are measured with these assays to demonstrate that infectious material exists. Fourth, other replication blocks are characterized as necessary.

Two recent publications affect the strategy herein. Jones and colleagues have recently reported the cloning of a human transcriptional elongation factor that interacts with tat (Cell 92:451-462, Feb. 20, 1998; A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA). The results of this work, presented at the March 1998 Keystone Symposium, showed that human Cyclin T interacts directly with tat in activating polII for elongation of messages driven by the HIV LTR. Jones transfected this gene into mouse cells and showed an increase in tat inducible gene expression. Introduction of this human gene into transgenic mice relieves one of the blocks to HIV replication.

The use of SCID-Hu mice for studying protease and RT inhibitors in vivo has been reported on (J. Infec. Dis. 177:337-346, 1998). HIV can replicate in this system and known RT and protease inhibitors inhibit replication. The broad use of SCID-Hu mice for drug studies is limited by the high cost of producing these mice which have to be individually repopulated with fetal human cells. Additionally, one will not be able to make use of genetic manipulation of the murine immune system, such as CD4 and CD8 knockouts, in this system. This study illustrates the utility of a mouse model for studying HIV. The approach herein has the potential to overcome the limitations of this model.

Evolution of whole virus

In one embodiment, the following steps are used to evolve HIV for replication in non-human cells. First, cloning vectors and protocols for shuffling infectious molecular clones in two non-infectious pieces are established. Second, methods for efficiently making large (>10^6 complexity) libraries of infectious molecules from shuffled fragments of HIV-1 are established. Third, libraries of HIV-1 recombinants are produced using in vivo recombination pathways. Fourth, synthetic

The library for is screened for replication in target cells by: Transfection of the library into human 293 cells, Coculture transfected cells with the target cells (monkey lymphocytes), Separate target cells from 293 cells, culture for 2 weeks, and passage the replicating virus. Because the amount of DNA used for transfection is limited (typically 30μg), the minimum amount of DNA required for producing one infectious virus determines the appropriate library size that can be analyzed in one transfection/infection. To determine the minimum infectious DNA dose, serial 10-fold dilution of the wild type HIV-1 DNA was used for transfection of human 293 cells in quintuplicate. Transfected cells were cocultured with human lymphoid cell line MT-4, to amplify infectious virus produced from transfected cells. Cultures were kept for 3 weeks to detect the end-point of infectivity. Approximately 10 ng DNA was required to produce one infectious virus. Therefore, libraries containing 3000 clones (30 μg divided by 10 ng) are adequate.

Based on this result, we decided to generate multiple sublibraries from the same assembly reaction. We made multiple aliquots from the assembly reaction, each of them containing 10^9 molecules. Presumably, all chimeric molecules from aliquot#1 should be different from any chimeric molecule in aliquot#2 (no redundancy). Each aliquot was amplified by PCR and cloned into the full-length HIV. Because cloning efficiency is not high, sizes of sublibraries are not 10^9, but range from 5,000 to 100,000. These are large enough because there is no advantage to make libraries larger than 3,000 (see the previous paragraph).

We next examined viability and diversity of one of the sublibraries. Six out of 40 randomly chosen clones were able to replicate in human MT-4 cells. When these clones were analyzed by Dra1 digestion, 13 clones exhibited patterns different from any one of the parental clones. Because Dra1 restriction digestion does not distinguish all parental clones (e.g. ELI, UG15, and Z2Z6 have the same restriction pattern therefore these three and chimeras between them are indistinguishable), recombination rate of 13/40 is very likely to be underestimation. This library has enough viability and diversity for screening.

Modifications can be made to the methods and compositions as herein before described without departing from the spirit or scope of the invention as claimed, and the invention can be put to a number of different uses. Assays kits or systems providing a use of any one of the components, methods or substrates herein before

WHAT IS CLAIMED IS:

1. A method for generating a viral polynucleotide sequence having a genotype encoding at least one modified viral phenotype, the method comprising:

contacting a cell or non-human animal which does not naturally support substantial replication of an predetermined virus, with at least one initial infectious virion or replicable genome of said predetermined virus under replication conditions;

recovering a plurality of replicated genome copies of said predetermined virus, either as virions or as viral genomes in polynucleotide form, wherein some or all of the replicated genome copies comprise a mutation relative to the initial infectious virion or replicable genome;

recombining a plurality of said replicated genome copies, so as to shuffle the mutations, thereby generating a collection of recombined replicated genome copies; and,

selecting or screening said collection of recombined replicated genome copies to obtain one or more replicable viral genome encoding at least one modified viral phenotype.

2. The method of claim 1, wherein the modified viral phenotype is a host range or cell tropism phenotype.

3. The method of claim 2, wherein the host range or cell tropism phenotype is the ability to replicate in mouse or macaque cells.

4. The method of claim 2, wherein the host range or cell tropism phenotype is the ability to replicate in a transgenic mouse expressing a human CD4 protein or HIV co-receptor on lymphocytes.

5. The method of claim 1, wherein the predetemined virus is selected from HIV-1, HIV-2, HCV, HBV and MLV.

6. The method of claim 5, wherein the virus is an HIV-1 which HIV-1 is a clinical isolate which has been passaged in cell culture for less than 10 passages.

16. The recombinant virus of claim 12, wherein the cell or organism is a transgenic mouse cell or a transgenic mouse, said transgenic cell or transgenic mouse harboring an expressible transgene encoding human CD4.

17. The recombinant virus of claim 16, wherein the transgenic cell or transgenic mouse further harbors an expressible transgene that encodes human CCR5.

18. A selected, shuffled virus having a genotype encoding at least one modified viral phenotype.

19. The selected, shuffled virus of claim 18, wherein said selected shuffled virus is an HIV-1 virus or a SHIV virus and replicates in a mouse cell.

20. The selected, shuffled virus of claim 19, wherein the mouse cell expresses human CD4 and human CCR5 encoded on a transgene or expression vector.

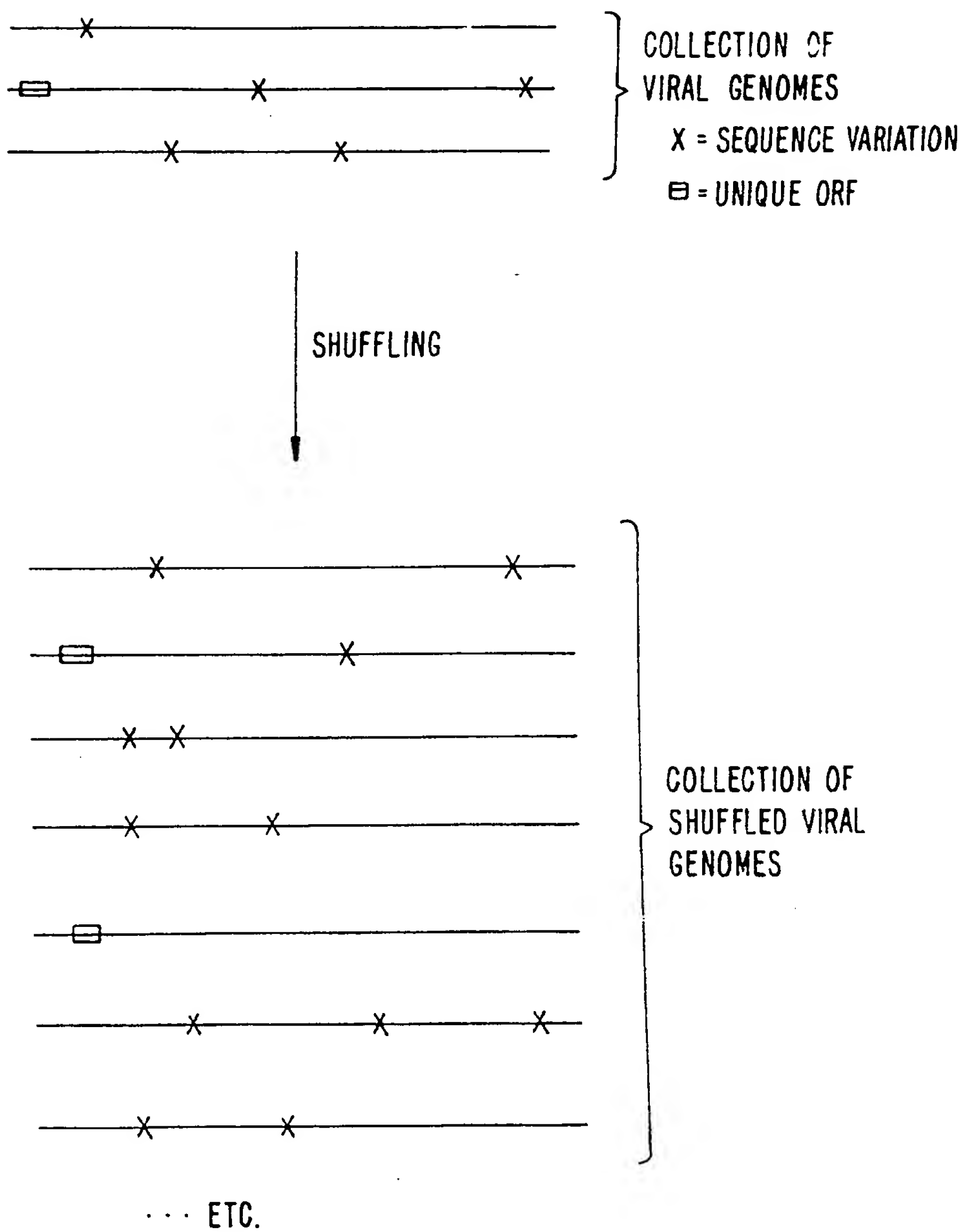
COLLECTION OF
VIRAL GENOMES
X = SEQUENCE VARIATION
= UNIQUE ORF
SHUFFLING
COLLECTION OF
SHUFFLED VIRAL
GENOMES
· · · ETC.

FIG. 2.

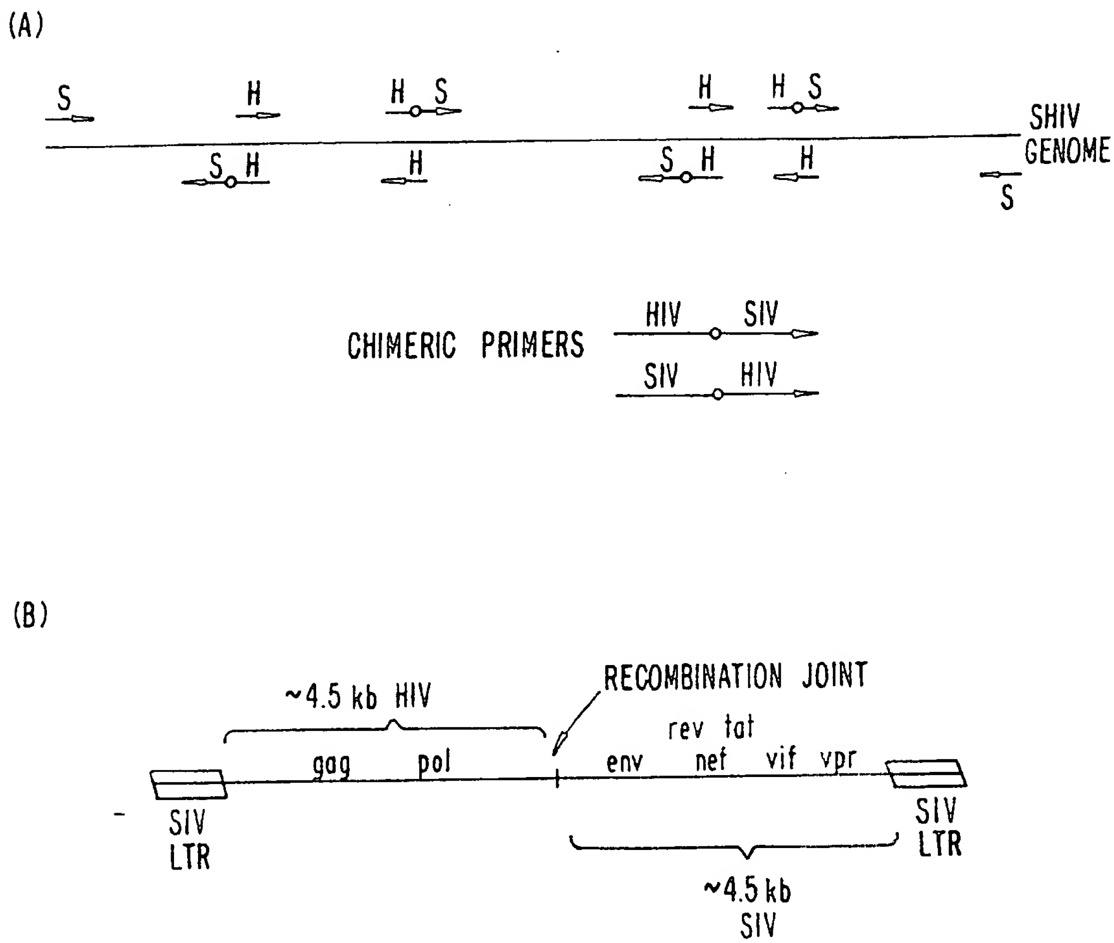
(A)
S
H
H S
H
H S
SHIV GENOME
S H
H
S H
H
S
CHIMERIC PRIMERS
HIV SIV
SIV HIV
(B)
~4.5 kb HIV
RECOMBINATION JOINT
rev tat
gag pol
env nef vif vpr
SIV LTR
SIV LTR
~4.5 kb SIV

FIG. 4.

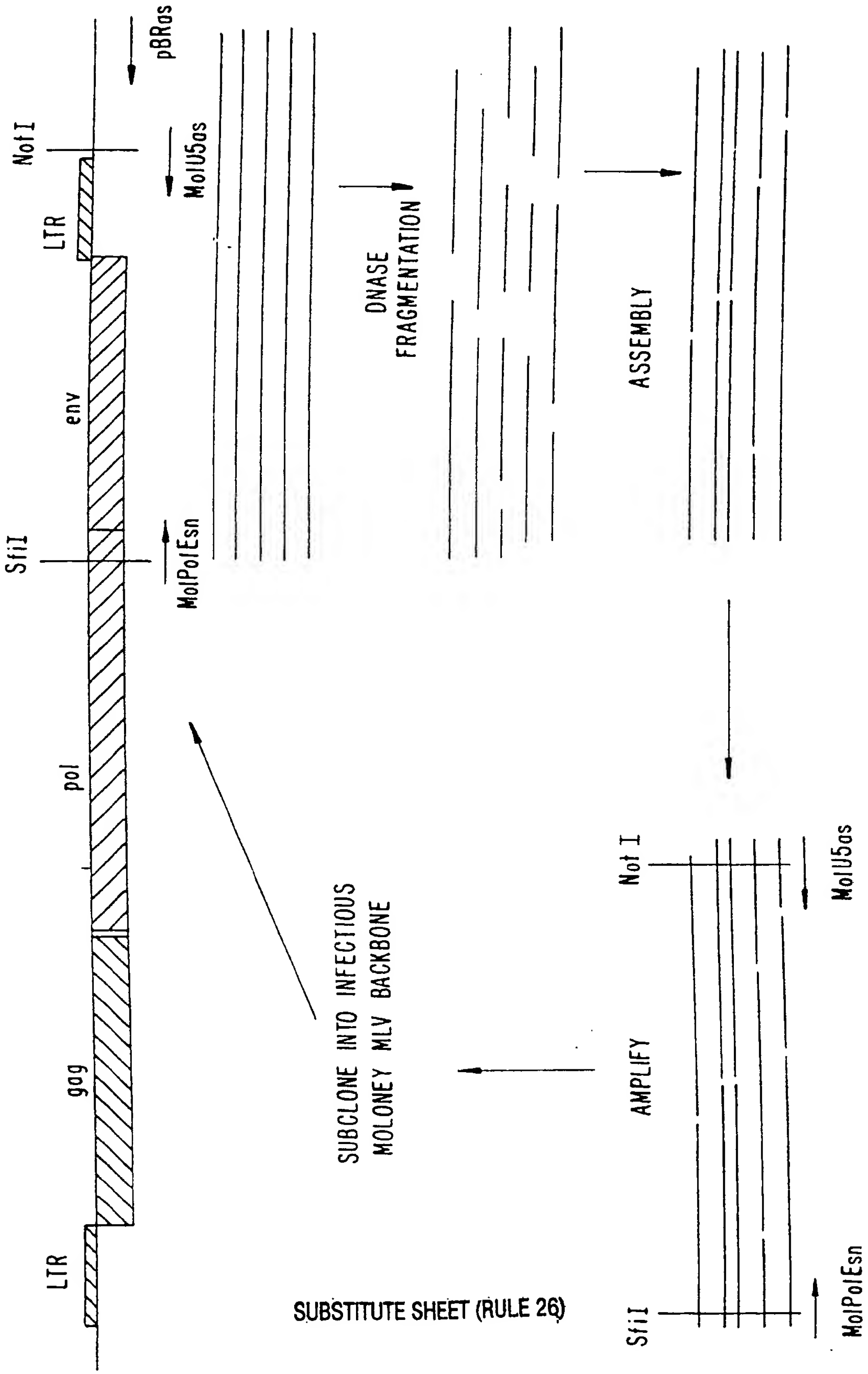
pBRas
Not I
MolU5as
LTR
env
SfiI
MolPolEsn
pol
gag
LTR
DNASE
FRAGMENTATION
ASSEMBLY
SUBCLONE INTO INFECTIOUS
MOLONEY MLV BACKBONE
AMPLIFY
Not I
MolU5as
SfiI
MolPolEsn

FIG. 6.

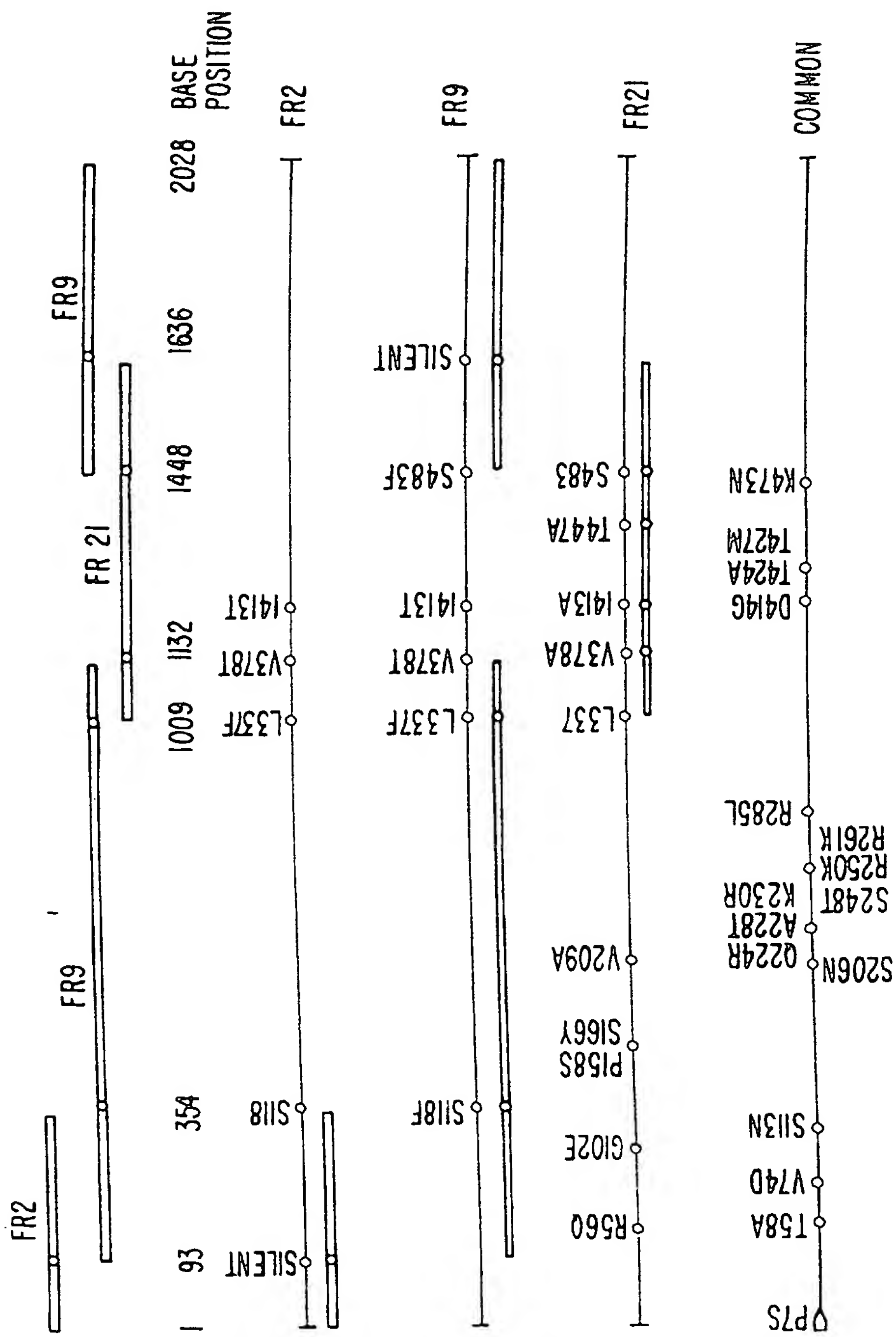
BASE POSITION
1
93
354
1009
1132
1448
1636
2028
FR2
SILENT
S118
L337F
V378T
I413T
FR9
S118F
L337F
V378T
I413T
S483F
SILENT
FR21
R56Q
G102E
P158S
S166Y
V209A
L337
V378A
I413A
T447A
S483
COMMON
P7S
T58A
V74D
S113N
S206N
Q224R
A228T
K230R
S248T
R250K
R261K
R285L
D419G
T424A
T427M
K473N
FR2
FR9
FR 21
FR9

FIG. 8.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/23107

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/5, 91.33, 235.1, 236, 239, 320.1; 424/199.1, 204.1, 205.1, 207.1, 208.1, 227.1, 228.1; 536/23.72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

U.S. Serial No. 08/962,236

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, WPIDS, USPATFUL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MAASSAB, H.F. et al. Characterization of an Influenza A Host Range Mutant. Virol. 1983, Vol. 130, pages 342-350, see entire document.	12, 13, 18
Y	PATTEN, P.A. et al. Applications of DNA shuffling to pharmaceuticals and vaccines. Curr. Opin. Biotech. 1997, Vol. 8, pages 724-733, see entire document.	1-11, 14-17, 19, 20
Y	STEMMER, W.P.C. Rapid evolution of a protein in vitro by DNA shuffling. Nature. 04 August 1994, Vol. 370, pages 389-391, see entire document.	1-11, 14-17, 19, 20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

12 FEBRUARY 1999

Date of mailing of the international search report

01 MAR 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer
D. Lawrence for
Jeffrey S. Parkin, Ph.D.

Telephone No. (703) 308-1234

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07H 21/04; C12Q 1/70; C12P 19/34; C12N 7/00, 7/04, 7/02, 15/00; A61K 39/12, 39/21, 39/29-

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/5, 91.33, 235.1, 236, 239, 320.1; 424/199.1, 204.1, 205.1, 207.1, 208.1, 227.1, 228.1; 536/23.72